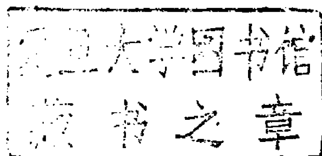
高等学校教材

随机过程引论

何声武





高等教育出版社

（京）112号

内容提要

本书分引论、离散时间的马尔可夫链、连续时间的马尔可夫链、更新过程等四章，只需要微积分和高等代数的预备知识即可阅读．可供需要掌握随机过程初步内容的高等学校有关专业作为教材使用，也可供研究生与工程技术人员参考．

本书由北京师范大学陈木法教授审稿．

图书在版编目（CIP）数据

随机过程引论/何声武．—北京：高等教育出版社，1999

ISBN 7－04－006804－4

Ⅰ．随…　Ⅱ．何…　Ⅲ．随机过程　Ⅳ．O211.6

中国版本图书馆 CIP 数据核字(98)第12374号

出版发行	高等教育出版社		
社　址	北京市东城区沙滩后街55号	邮政编码	100009
电　话	010—64054588	传　真	010—64014048
网　址	http://www.hep.edu.cn		
经　销	新华书店北京发行所		
印　刷	北京华文印刷厂		
开　本	787×1092　1/16	版　次	1999年6月第1版
印　张	13.5	印　次	1999年6月第1次印刷
字　数	330 000	定　价	13.50元

凡购买高等教育出版社图书，如有缺页、倒页、脱页等质量问题，请在所购图书销售部门联系调换。

序　言

本书是为数学专业(包括概率统计专业与应用数学专业)的大学生、研究生学习随机过程课程而编写的入门教材,这门课程应是初等概率论的后继课程,要求学生有微积分与高等代数的数学准备,不需要测度论、泛函分析等较高深的知识.本书内容包括马尔可夫链与更新过程,它们是应用概率论的最重要的理论基础部分,因此对于希望了解与掌握随机过程的初步内容与应用方法的各类理工科、医科等大学生与研究生以及科学技术人员,本书也是适用的.

实践表明,即使对概率论与随机过程方向的研究生,一开始就学以测度论为基础的随机过程课程,学起来也有一定的困难.从初等的、直观意义比较明显的随机过程学起是较好的选择.对于以应用为目的而学习随机过程的人,自然就更需要随机过程的初等入门教材.尽管目前国内外出版的随机过程的教材已经不少,然而读者可读、教师可教的教材仍难觅得.作者在多年的教学工作中为编写一部实用的随机过程教材煞费苦心,孜孜以求.首次的尝试就是1989年出版的《随机过程导论》一书.该书中写得比较成功,最受欢迎的只是其中的马尔可夫链部分.不用测度论与泛函分析,确实给讲授与学习随机过程造成很大的困难.本书只包含马尔可夫链与更新过程,不得不舍去了布朗运动等重要内容的原因就在于此.本书中的马尔可夫链部分即是在该书的基础上修改而成的.选取更新过程,主要是由于它应用的广泛性,也考虑到可以用较直观的条件数学期望讲述它的理论.这部分内容也是在作者多年使用的讲义基础上写成的.

与概率论一样,随机过程有着广泛的直观背景,因此我们十分强调解释结果的直观意义,并且给出了众多的应用例子.这是讲述随机过程理论十分有利的一面.考虑到数学专业的学生使用的需要,与一般的数学教科书一样,本书遵循数学理论的严谨性原则.选讲的内容在理论上是完整的、封闭的,包含了全部主要的结论及其证明.这有利于让学生了解所讨论问题的全貌及实质,学会系统地解决问题的思路与方法.事实上,许多应用工作者也不欢迎那种内容显得支离破碎,证明经常略去,只是堆砌一大堆定义与定理的随机过程教材.这也许是不少声称面向应用的随机过程教材不太受欢迎的一个重要原因.为应用目的而选用本书的读者可以放心地按照自己的实际情况,略过理论性较强的内容及证明.尽管本书主要是为课堂教学编写的,由于论述与说明比较详细,也适合于自学.事实上,并非书中所有的内容都必须在课堂上讲解.教师只需讲授最重要的结论与例子,余下部分留作参考,或让学生自行阅读.

为了使学生一开始就对随机过程有具体的了解,我们在作为引论的第一章中就讲述了最基本的两类随机过程:伯努利过程与泊松过程.第二章是离散时间马尔可夫链的初等理论.在已知转移概率矩阵的条件下,对马尔可夫链作全面的分析,包括状态的分类,状态性质的判别,求吸收概率、平均吸收时间和平稳分布等等.学习这一章只需有微积分的准备即可.在本章中比较充分地展开与运用了我国概率论学者惯用的非负系数线性方程组的最小解方法.第三章连续时间马尔可夫链集中讲述规则链的理论.作这样的选择,一方面是因为可以用初等方法完整地展开理论,另一方面对于应用来说它已经足够了,因为在实际应用中遇到的都是规则链.第四章是更新理论.我们强调了更新技巧的运用,这个方法是十分有用的,也是比较容易学会的.为了使本书的

内容尽可能地接近随机过程的现代理论,我们引进了停时的概念,但不作深入的展开;在马尔可夫链的转移概率的极限定理与更新定理的证明中使用了耦合方法,避免了很难看出概率意义的纯分析的证明,但这也只是给与读者概率论方法的熏陶,并非要求读者立刻学会使用这个方法.掌握马尔可夫链及更新过程的基本理论足以应付应用随机过程的初步需要.书末附有中文参考书目,列入适合本书读者参考的不用测度论的随机过程的教材,没有收入要用测度论的随机过程的书籍.

本书的编写得到了国内同行的热情鼓励与支持,得益于汪嘉冈教授的随机过程讲义,在此一并致谢.一套可读可教的、从初等入门到现代前沿水平的、系列的随机过程的教材是作者多年来所梦寐以求的.期望,也许是奢望本书成为贡献给这一宏大工程的一块垫脚石.

作　者

1997 年 7 月于上海

目　录

第一章 引 论

§1.1 什么是随机过程

现实世界中的许多现象是随时间的进展而变化与发展的,这些现象通常即称之为过程.例如,质点的运动过程就是最普通的过程的例子.假设一个质点在直线上运动,它的位置随时间而改变.数学上,质点的位置可表示为时间 t 的函数 $X(t)$.通常假定质点是在一个确定的力的作用下运动的,则表示运动过程的函数 $X(t)$ 也是完全确定的.如果质点是在一个随机的力(它由各种偶然因素所形成)的作用下,那么质点的运动也是随机的.换句话说,在时刻 t 质点的位置 $X(t)$ 不再是一个确定的数,而是一个随机变量.这样,我们就得到一族无穷多个随机变量 $\{X(t),t\geqslant 0\}$(取初始时刻为 $t=0$),它描述了随机的运动过程.这样的例子还可以举出很多.例如,在一条生产线上,我们对产品逐个进行检查,以 $N(t)$ 表示一天从开工($t=0$)到时刻 t 累计的次品数,或者在一个电话交换站里,以 $N(t)$ 表示一天从上班($t=0$)到时刻 t 为止接到的呼叫次数,那么 $\{N(t),0\leqslant t\leqslant a\}$ 就是一族随时间而变化的随机变量,它描述了次品数或呼叫数累积的过程.再如在保险公司里,以 $R(n)$ 表示从年初到第 n 天里累计的因火灾而付出的赔偿.在气象观测站里,以 $T(n)$ 表示一年中第 n 天的平均温度.$\{R(n),1\leqslant n\leqslant N\}$ 或 $\{T(n),1\leqslant n\leqslant N\}$ 都是一族随时间的进展而变化的随机变量,只是这里的记录不是按时间连续地进行的,而是按一个个时间单元而记录的.所以,我们通常得到的是一列随机变量,它描述了赔偿的累计过程或气温的变化过程.总之,随时间的进展而变化与发展的随机现象就自然地被称为随机过程.

现在给出随机过程的数学定义.

定义 设对每个 $t\in T$,$X(t)$ 为一个随机变量,则随机变量族 $X=\{X(t),t\in T\}$ 称为随机过程,t 称为 X 的指标或参数,T 为指标集或参数集.

随机过程的指标集 T 可以是任意的集合,但通常是数直线的某个子集,即指标代表时间,在本课程里就是如此.当 T 为可列集 $\mathbb{N}=\{0,1,2\cdots\}$ 或 $\mathbb{Z}=\{\cdots,-1,0,1,\cdots\}$ 时,随机过程 X 称为离散时间(参数)的.离散时间的随机过程也称为随机序列.若 $T=[a,b]$ 时(a,b 可以是有限或无限的实数),称随机过程 X 为连续时间(参数)的.

$X=\{X(t),t\in T\}$ 这族随机变量,可能取到的值称为随机过程的状态,状态的全体 E 称为随机过程 X 的状态空间或相空间.当 E 为可列集 $\{x_0,x_1,\cdots,x_n,\cdots\}$ 时,称 X 为离散状态的.这时每个随机变量 $X(t)$ 都是离散型随机变量.当 X 不是离散状态时,则称 X 为连续状态的.

当我们对随机过程 $X=\{X(t),t\in T\}$ 作观察时,对每个 t 得到 $X(t)$ 的一个观察值,因而得到的是 t 的一个函数,这个函数称为 X 的一个样本函数或一条轨道.若作多次观察,由于随机性,可能得到不同的样本函数.换句话说,观察到什么样的样本函数是随机的.因此很自然地,随机过程也称为随机函数.随机函数 X 在 t 点的值就是随机变量 $X(t)$.

随机过程的指标集 T 也可以是两维或维数更高的集合.例如在研究海洋波浪时,将某海域看作一个平面区域 D,以 $X(x,y,t)$ 记时刻 t 在坐标为 (x,y) 的点处的海浪高度,那么 $X=$

$\{X(x,y,t),(x,y)\in D,t\geqslant 0\}$是有三维指标集的随机过程,指标集 T 为$D\times[0,\infty)$.指标集是两维或维数更高的集合的随机过程称为多指标随机过程或随机场(把指标向量解释为空间中点的坐标向量).本课程将不涉及多指标随机过程.

同样地,随机过程 X 的状态空间E 也可以是d 维空间 $\mathbf{R}^d$ 的子集,即对每个 t,$X(t)=(X_1(t),X_2(t),\cdots,X_d(t))$是 d 维随机变量.这时 $X=\{(X_1(t),X_2(t),\cdots,X_d(t)),t\in T\}$称为 d 维随机过程,它的每个分量是一维随机过程 $X_k=\{X_k(t),t\in T\}$,$k=1,2,\cdots,d$.除非特别指明,本课程只讨论一维随机过程.

在概率论课程中我们知道,一个随机变量的概率统计特性由它的分布函数决定,多维随机变量的概率统计特性由它的联合分布函数决定.由于随机过程实际上是无穷多个随机变量,因此必须用全部可能的有限维联合分布函数来决定它的概率统计特性.对任意的正整数 $n\geqslant 1$,任取指标集 T 中n 个不同的指标$t_1,t_2,\cdots,t_n$,称

$$F_{t_1,t_2,\cdots,t_n}(x_1,x_2,\cdots,x_n)=\mathrm{P}(X(t_1)\leqslant x_1,X(t_2)\leqslant x_2,\cdots,X(t_n)\leqslant x_n)$$

为随机过程 X 的一个*有限维*(n 维)*分布*(注意:我们使用右连续的分布函数).变化 n 及t_1,$t_2,\cdots,t_n$ 所得到的有限维分布全体称为随机过程X 的*有限维分布族*.随机过程的分布就是指它的有限维分布族.

在概率论课程中我们也知道,可以用随机变量的数字特征来描述随机变量的概率统计特性,如数学期望、方差与协方差等.对于随机过程,也同样可以讨论它的数字特征.

若所有 $X(t)$的数学期望都存在,则称

$$m(t)=\mathrm{E}[X(t)],\quad t\in T$$

为随机过程 X 的*均值函数*.若所有 $X(t)$的方差都存在,则称

$$C(t,s)=\mathrm{E}[(X(t)-m(t))(X(s)-m(s))],\quad t,s\in T$$

为随机过程 X 的*协方差函数*.

例 1.1　正态过程　随机过程 $X=\{X(t),t\in T\}$称为*正态过程*,若它的一切有限维分布是正态分布.与正态随机变量相似,正态过程的分布由它的均值函数及协方差函数完全决定,正态过程也是实际应用中最常遇到的随机过程之一.　□

例 1.2　平稳过程　随机过程 $X=\{X(t),t\geqslant 0\}$称为(*严格*)*平稳过程*,若对一切 $n\geqslant 1$,$t_1<t_2<\cdots<t_n$,$t>0$,$(X(t_1),X(t_2),\cdots,X(t_n))$的分布与$(X(t_1+t),X(t_2+t),\cdots,X(t_n+t))$的分布相同,即 X 一切有限维分布对时间的推移保持不变.这时可取任一时刻 s 作为初始时刻,而过程 $X^{(s)}=\{X_{t+s},t\geqslant 0\}$的分布与 X 的分布仍然相同.特别,X 的一维分布,即 $X(t)$的分布与时间 t 无关;$(X(t),X(s))$的二维分布只依赖于 $t-s$.

从数学讨论方便的角度出发,平稳随机过程的指标集应更合理地取为$(-\infty,+\infty)$或全体整数(离散时间情形),因为这时时间的推移既可以是正的,也可以是负的.然而从实际应用的角度看,要求追溯到无穷的过去有点不太现实.　□

例 1.3　独立增量过程　随机过程 $X=\{X(t),t\geqslant 0\}$称为*独立增量过程*,若对一切 $n\geqslant 1$,$0\leqslant t_0<t_1<\cdots<t_n$,

$$X(t_0), X(t_1)-X(t_0), X(t_2)-X(t_1), \cdots, X(t_n)-X(t_{n-1})$$

相互独立.若对一切$0\leqslant s<t$,增量$X(t)-X(s)$的分布只依赖于$t-s$,则称X有平稳增量.有平稳增量的独立增量过程简称为独立平稳增量过程.

若$X=\{X(n), n\geqslant 0\}$是离散时间的独立增量过程,令

$$\xi(0)=X(0), \quad \xi(n)=X(n)-X(n-1), n\geqslant 1,$$

则$\{\xi(n), n\geqslant 0\}$是一列独立随机变量.因此,离散时间的独立增量过程就是独立随机变量部分和.不难看出,这时X有平稳增量等价于$\{\xi(n), n\geqslant 1\}$同分布.独立同分布一般简写为i.i.d.. □

为了让初学的读者先接触一点随机过程的具体材料,在这章引论的下两节中,我们分别介绍两个常见而又简单的随机过程:伯努利(Bernoulli)过程与泊松(Poisson)过程.

习　　题

1-1. 设ξ为一非负随机变量,分布函数为F.令

$$X_t=\min(\xi, t), \quad t\geqslant 0.$$

求$X=\{X_t, t\geqslant 0\}$的有限维分布.

1-2. 设$C(t,s)$为随机过程$X=\{X(t), t\in T\}$的协方差函数.证明:

(1) 对称性:$C(t,s)=C(s,t), \quad t,s\in T$;

(2) 非负定性:对任意正整数$n\geqslant 1, t_1, t_2, \cdots, t_n$及实数$a_1, a_2, \cdots, a_n$,

$$\sum_{i,j=1}^{n} a_i a_j C(t_i, t_j)\geqslant 0.$$

1-3. 正态过程$X=\{X(t), t\geqslant 0\}$是平稳过程的充要条件为X的均值函数$m(t)$是常数,协方差函数$C(t,s)=R(t-s)$只依赖于$t-s$.

1-4. 设$X=\{X(t), t\geqslant 0\}$为零均值平稳独立增量正态过程,且$X(0)=0$.写出X的有限维分布族.

§1.2　伯努利过程

定义　随机过程$X=\{X_n, n\geqslant 1\}$称为伯努利过程,若$\{X_n, n\geqslant 1\}$为i.i.d.随机变量序列,且

$$\mathrm{P}(X_n=1)=p, \ \mathrm{P}(X_n=0)=q=1-p, \ 0<p<1, \ n\geqslant 1.$$

伯努利过程$X=\{X_n, n\geqslant 1\}$的状态空间为$\{0,1\}$,指标集为$\{1,2,\cdots\}$.伯努利过程是最简单,又最常见的随机过程.例如,在装配线上对产品进行检验.令

$$X_n=\begin{cases}1, & \text{第 } n \text{ 件是次品},\\ 0, & \text{第 } n \text{ 件是好品}.\end{cases}$$

当各件产品是好是次互不相关时,$\{X_n\}$就是伯努利过程.又如对通过十字路口的车辆进行观测,令

$$X_n = \begin{cases} 0, \text{第 } n \text{ 部车辆直走}, \\ 1, \text{第 } n \text{ 部车辆转弯}, \end{cases}$$

则$\{X_n, n \geqslant 1\}$也可被假定是一个伯努利过程.伯努利过程描述了一列独立重复试验,通常把事件$\{X_n = 1\}$称为第 n 次试验成功,把事件$\{X_n = 0\}$称为第 n 次试验失败.每次试验成功的概率为 p.各次试验相互独立.

设 $X = \{X_n, n \geqslant 1\}$为伯努利过程.令

$$N_0 = 0, \quad N_n = \sum_{j=1}^{n} X_j, \quad n \geqslant 1. \tag{1}$$

则在概率论中熟知,$n \geqslant 1$ 时

$$\mathsf{P}(N_n = k) = \mathrm{C}_n^k p^k q^{n-k}, \qquad 0 \leqslant k \leqslant n,$$
$$\mathsf{E}[N_n] = np, \quad \mathrm{Var}[N_n] = npq.$$

事实上,$N = \{N_n, n \in N\}$是平稳独立增过程.增量 $N_{n_2} - N_{n_1}(n_1 < n_2)$与 $N_{n_2 - n_1}$同分布.

定理 2.1 设$\{X_n, n \geqslant 1\}$为 i.i.d.随机变量序列,$\mathsf{E}|X_1| < \infty$,r 为一取正整数值的随机变量,$\mathsf{E}\, r < \infty$,且对一切 $n \geqslant 1$,$\{r = n\}$与$\{X_{n+1}, X_{n+2}, \cdots\}$相互独立,则$\mathsf{E}\left|\sum_{n=1}^{r} X_n\right| < \infty$,且

$$\mathsf{E}\{X_1 + X_2 + \cdots + X_r\} = \mathsf{E}\, r\ \mathsf{E}\, X_1. \tag{2}$$

证 先设 $X_n \geqslant 0, n \geqslant 1$.

$$\mathsf{E}\{X_1 + \cdots + X_r\} = \sum_{k=1}^{\infty} \sum_{n=1}^{k} \mathsf{E}\{X_n 1_{\{r=k\}}\} = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathsf{E}\{X_n 1_{\{r=k\}}\}$$
$$= \sum_{n=1}^{\infty} \mathsf{E}\{X_n 1_{\{r \geqslant n\}}\} = \sum_{n=1}^{\infty} \mathsf{E}\, X_n \mathsf{P}(r \geqslant n) = \mathsf{E}\, r\ \mathsf{E}\, X_1.$$

这里我们用 1_A 记事件 A 的示性随机变量:

$$1_A = \begin{cases} 1, \text{若事件 } A \text{ 发生}, \\ 0, \text{若事件 } A \text{ 不发生}. \end{cases}$$

由于上面的级数是正项级数,因此两个求和号可以交换.又由于$\{r \geqslant n\}$是$\{r < n-1\}$的逆事件,因而由假设与 X_n 独立,故得第四个等式.

将上面证得的结果用于$\{X_n^+ = \max(X_n, 0), n \geqslant 1\}$及$\{X_n^- = \max(-X_n, 0), n \geqslant 1\}$可得

$$\mathsf{E}\{X_1^+ + \cdots + X_r^+\} = \mathsf{E}\, r\ \mathsf{E}\, X_1^+,$$
$$\mathsf{E}\{X_1^- + \cdots + X_r^-\} = \mathsf{E}\, r\ \mathsf{E}\, X_1^-.$$

将上两式相减即得(2),因为 $X_n = X_n^+ - X_n^-$. □

(2)式通常称为瓦尔德(Wald)等式.它解决了求一类随机个 i.i.d.随机变量和的数学期望的问题,是一个有力的研究工具.在定理 2.1 中,若又有 $X_1 \geqslant 0$,由证明可看出,没有条件$\mathsf{E}\, r < \infty$时(2)式仍成立,且$\mathsf{E}\left[\sum_{n=1}^{r} X_n\right] < \infty$当且仅当$\mathsf{E}\, r < \infty$.

定义 设给定随机序列 $X = \{X_n, n \geqslant 1\}$.$r$ 为一个取正整数值(包括$+\infty$)的随机变量.若对

一切 $n\geqslant 1$,事件$\{\tau=n\}$是否发生由$(X_1,\cdots,X_n)$取的值确定,则称 τ 为 X 的停时.停时是停止时间的简称.

推论 设 $X=\{X_n,n\geqslant 1\}$为 i.i.d. 随机变量序列,$\mathsf{E}|X_1|<\infty$,τ 为 X 的停时,$\mathsf{E}\,\tau<\infty$,则 $\mathsf{E}|\sum_{n=1}^{\tau}X_n|<\infty$,且

$$\mathsf{E}\{X_1+X_2+\cdots+X_\tau\}=\mathsf{E}\,\tau\ \mathsf{E}\,X_1.$$

证 只需注意,对一切 $n\geqslant 1$,事件$\{\tau=n\}$是否发生由$(X_1,\cdots,X_n)$取的值确定,从而与$\{X_{n+1},X_{n+2},\cdots\}$独立,满足定理 2.1 的假设. □

停时是近代概率论与随机过程理论的重要概念之一,也是经常遇到的研究对象.例如,设 $X=\{X_n,n\geqslant 1\}$为伯努利过程,则首次成功的时刻

$$T_1=\inf\{n\geqslant 1:X_n=1\} \tag{3}$$

就是 X 的停时,因为

$$\{T_1=1\}=\{X_1=1\},\ \{T_1=n\}=\{X_1=0,\cdots,X_{n-1}=0,X_n=1\},\ n\geqslant 2.$$

我们约定,空集的下确界为 $+\infty$.这是合理的.例如在定义(3)中,$\{n\geqslant 1:X_n=1\}$是空集表示试验一直失败,首次成功的时刻自然应定义为 $+\infty$.更一般地,第 k 次成功的时刻

$$T_k=\inf\{n\geqslant 1:N_n=k\},\quad k\geqslant 1 \tag{4}$$

也是停时,因为

$$\{T_k=k\}=\{N_k=k\},\ \{T_k=n\}=\{N_1<k,\cdots,N_{n-1}<k,N_n=k\},\ n\geqslant k+1,$$

而 $N_1,\cdots,N_n$ 的值完全由 $X_1,\cdots,X_n$ 的值决定.停时是一个富有概率论特色的概念,它的引入大大加深了随机过程理论的研究.下面的定理就是一个这样的例子.

定理 2.2 设 $X=\{X_n,n\geqslant 1\}$为 i.i.d. 随机变量序列,τ 为 X 的停时,且 $\mathsf{P}(\tau<\infty)=1$.令

$$Y_n=X_{\tau+n},\quad n\geqslant 1,$$

则 $Y=\{Y_n,n\geqslant 1\}$与 $X=\{X_n,n\geqslant 1\}$同分布,且与$(\tau,X_1,\cdots,X_\tau)$相互独立.

证 假设 X 的分布函数为 F,对任意 $n,m\geqslant 1$ 及实数 $x_1,\cdots,x_n,y_1,\cdots,y_m$,

$$\begin{aligned}
&\mathsf{P}(\{X_1\leqslant x_1,\cdots,X_n\leqslant x_n,\tau=n\}\cap\{Y_1\leqslant y_1,\cdots,Y_m\leqslant y_m\})\\
&=\mathsf{P}(\{X_1\leqslant x_1,\cdots,X_n\leqslant x_n,\tau=n\}\cap\{X_{n+1}\leqslant y_1,\cdots,X_{n+m}\leqslant y_m\})\\
&=\mathsf{P}(X_1\leqslant x_1,\cdots,X_n\leqslant x_n,\tau=n)\mathsf{P}(X_{n+1}\leqslant y_1,\cdots,X_{n+m}\leqslant y_m)\\
&=\mathsf{P}(X_1\leqslant x_1,\cdots,X_n\leqslant x_n,\tau=n)F(y_1)\cdots F(y_m).
\end{aligned} \tag{5}$$

在(5)式中令所有的 $x_1,\cdots,x_n$ 都趋于正无穷,然后对所有的 $n\geqslant 1$ 相加可得

$$\mathsf{P}(Y_1\leqslant y_1,\cdots,Y_m\leqslant y_m)=F(y_1)\cdots F(y_m).$$

这即表明 $Y=\{Y_n,n\geqslant 1\}$为 i.i.d. 序列,且与 X 同分布.现在(5)式就是

$$\mathsf{P}(\{X_1\leqslant x_1,\cdots,X_n\leqslant x_n,\tau=n\}\cap\{Y_1\leqslant y_1,\cdots,Y_m\leqslant y_m\})$$

$$=\mathrm{P}(X_1\leqslant x_1,\cdots,X_n\leqslant x_n,r=n)\mathrm{P}(Y_1\leqslant y_1,\cdots,Y_m\leqslant y_m).$$

这即表明 Y 与 $r,X_1,\cdots,X_r$ 相互独立. □

设 $\{X_n,n\geqslant1\}$ 为 i.i.d. 随机变量序列. 我们原来只知道, 对固定的时刻 n, $\{X_1,\cdots,X_n\}$ 与 $\{X_{n+1},X_{n+2},\cdots\}$ 相互独立, 且 $\{X_{n+1},X_{n+2},\cdots\}$ 与 $\{X_1,X_2,\cdots\}$ 同分布. 定理 2.2 告诉我们, 对 X 的有限停时 r 这些性质仍然保持, 特别, $\{X_{r+1},X_{r+2},\cdots\}$ 与 $\{X_1,X_2,\cdots\}$ 同分布表明, 在 r 之后, 随机序列又恢复原来的概率统计特性. 例如, 对于伯努利过程, 在得到首次成功之后, 就好像独立重复试验又从头开始一样. 直观上, 人们很容易接受这个结论. 但是首次成功的时刻是随机的时刻, 这个结论并不显然, 并非在任何的随机时刻之后, 独立重复试验都与从头开始一样.

若停时 r 可能取 $+\infty$ 值, 那么从(5)式我们得到的是

$$\mathrm{P}(r<\infty,Y_1\leqslant y_1,\cdots,Y_m\leqslant y_m)=\mathrm{P}(r<\infty)F(y_1)\cdots F(y_m),$$
$$\mathrm{P}(Y_1\leqslant y_1,\cdots,Y_m\leqslant y_m\mid r<\infty)=F(y_1)\cdots F(y_m),$$

即在 $r<\infty$ 的条件下, $Y=\{Y_1,\cdots,Y_n,\cdots\}$ 的条件分布与 X 的分布相同. 事实上, Y 只有在 $\{r<\infty\}$ 的条件下才有定义.

定理 2.3 设 $X=\{X_n,n\geqslant1\}$ 为伯努利过程, T_k 为第 k 次成功的时刻(由(4)定义). 令

$$W_1=T_1,\quad W_n=T_n-T_{n-1},\quad n\geqslant2, \tag{6}$$

则 $\{W_n,n\geqslant1\}$ 为 i.i.d. 随机变量序列, 且共同分布为几何分布:

$$\mathrm{P}(W_n=k)=pq^{k-1},\quad k\geqslant1. \tag{7}$$

T_n 服从帕斯卡(Pascal)分布:

$$\mathrm{P}(T_n=k)=\mathrm{C}_{k-1}^{n-1}p^kq^{k-n},\quad k\geqslant n, \tag{8}$$

$$\mathrm{E}\,T_n=\frac{n}{p},\quad \mathrm{Var}\,T_n=\frac{nq}{p^2}. \tag{9}$$

证 我们熟知, 首次成功时刻 $W_1=T_1$ 服从几何分布:

$$\mathrm{P}(W_1=k)=pq^{k-1},\quad k\geqslant1.$$

不难看出, W_2 是 $X^{(2)}=\{X_{T_1+n},n\geqslant1\}$ 的首次成功时刻; $W_2+\cdots+W_k,k\geqslant2$ 是 $X^{(2)}$ 的第 $k-1$ 次成功时刻. 由定理 2.2, W_1 与 $X^{(2)}$ 独立, $X^{(2)}$ 与 X 同分布, 因此 W_2 也服从几何分布(7), 且 W_1 与 $\{W_2,W_3,\cdots\}$ 相互独立. 一般地, 对 $n\geqslant2$, W_n 是 $X^{(n)}=\{X_{T_{n-1}+k},k\geqslant1\}$ 的首次成功时刻; $W_n+\cdots+W_k,k\geqslant n$ 是 $X^{(n)}$ 的第 $k-n+1$ 次成功时刻. 由定理 2.2, $(W_1,\cdots,W_{n-1})$ 与 $X^{(n)}$ 独立, $X^{(n)}$ 与 X 同分布, 因此 W_n 也服从几何分布(7), 且 $(W_1,\cdots,W_{n-1})$ 与 $\{W_n,W_{n+1},\cdots\}$ 相互独立. 总之, $\{W_n,n\geqslant1\}$ 为独立同几何分布(7)的随机变量序列. 最后, T_n 是 n 个独立同几何分布随机变量之和, 故服从帕斯卡分布(8). 又 $\mathrm{E}\,W_1=1/p$, $\mathrm{Var}\,W_1=q/p^2$, 故(9)成立. □

定理 2.3 提供了, 对伯努利过程 $X=\{X_n,n\geqslant1\}$, 由(1)定义的平稳独立增量过程 N 的新的构造: 从一列独立同几何分布(7)的随机变量 $\{W_n,n\geqslant1\}$ 出发, 作 $T_n=W_1+\cdots+W_n,n\geqslant1$ 及

$$N_n=\begin{cases}0, & n<T_1,\\ k, & T_k\leqslant n<T_{k+1},\ k\geqslant 1.\end{cases}\tag{10}$$

$\{W_n\}$通常称为伯努利过程的等待时间序列，因为 W_n 是等待第 n 次成功出现的时间．我们知道，几何分布具有无后效性，即对任意非负整数 k,m，

$$\mathsf{P}(W_n>k+m\mid W_n>k)=\mathsf{P}(W_n>m).\tag{11}$$

这无后效性与伯努利过程的独立同分布性相联系．事实上不难证明，一个只取正整数值的随机变量若具有无后效性(11)，其分布必为几何分布．

定理 2.4 设 T_k 为伯努利过程 $X=\{X_n,n\geqslant 1\}$ 的第 k 次成功的时刻，$k\geqslant 1$，数列 $\{f(n),n\geqslant 1\}$ 满足 $\sum_{n=1}^{\infty}|f(n)|<\infty$，则

$$\mathsf{E}\Big[\sum_{k=1}^{\infty}f(T_k)\Big]=p\sum_{n=1}^{\infty}f(n).\tag{12}$$

证 注意到

$$f(T_1)=f(1)X_1+\cdots+f(T_1-1)X_{T_1-1}+f(T_1)X_{T_1},$$

$$\cdots\cdots$$

$$f(T_k)=f(T_{k-1}+1)X_{T_{k-1}+1}+\cdots+f(T_k-1)X_{T_k-1}+f(T_k)X_{T_k},$$

故

$$\mathsf{E}\Big[\sum_{k=1}^{\infty}f(T_k)\Big]=\mathsf{E}\Big[\sum_{n=1}^{\infty}f(n)X_n\Big]=\sum_{n=1}^{\infty}f(n)\mathsf{E}[X_n]=p\sum_{n=1}^{\infty}f(n).\quad\square$$

例 2.1 设在每一时间单元内，一仪器内某元件损坏与否构成一个伯努利过程，损坏后替换一个新的元件耗费 c 元．为使仪器有元件损坏时，用户能享受免费调换，厂方在卖出产品时多收些费用，留作保修费．考虑到，现在投资 1 元，到下一个时间单元可增值为 $1/\alpha$ 元($\alpha<1$)．问为保证该仪器长期使用，厂方在卖出产品时，平均应多收多少费用？

解 以 N_n 表示到时刻 n 损坏的元件总数．T_k 由(4)定义．在时刻 T_k，第 k 次损坏了一个元件．为调换这一元件，开始售出时应多收 $c\alpha^{T_k}$元(到时刻 T_k 这些钱恰增值为 c 元)．因此，调换全部损坏的元件应多收 $D=\sum_{k=1}^{\infty}c\alpha^{T_k}$元，利用定理 2.4，其平均值为

$$\mathsf{E}\,D=\sum_{k=1}^{\infty}\mathsf{E}\,(c\alpha^{T_k})=cp\sum_{k=1}^{\infty}\alpha^k=c\,\frac{\alpha p}{1-\alpha}.\quad\square$$

习　题

2-1．设给定随机序列 $X=\{X_n,n\geqslant 1\}$．r_1,r_2 为 X 的两个停时．证明：

$$\max(r_1,r_2),\quad \min(r_1,r_2),\quad r_1+r_2$$

均为停时．

2-2．设 $X=\{X_n,n\geqslant 1\}$为一列独立同(0,1)上均匀分布的随机变量，$d\in(0,1)$，则

$$T=\inf\{n\geqslant 1:X_n\geqslant d\}$$

为 X 的停时,并求 $\mathsf{E}\,T,\mathsf{E}\,X_T$.

2-3. 设 $X=\{X_n,n\geqslant 1\}$ 为一列独立同(0,1)上均匀分布的随机变量,则

$$T=\inf\{n\geqslant 1:X_1+\cdots+X_n\geqslant 1\}$$

为 X 的停时,并证明 $\mathsf{E}\,T=\mathrm{e}$.

2-4. 设$\{X_n,n\geqslant 1\}$为 i.i.d.随机变量序列,$\mathsf{E}\,X_1=0$,$\mathsf{E}\,X_1^2<\infty$,r 为 X 的停时,$\mathsf{E}\,r<\infty$,则

$$\mathsf{E}\,\{X_1+X_2+\cdots+X_r\}^2=\mathsf{E}\,r\;\mathsf{E}\,X_1^2.$$

2-5. 设$\{X_n,n\geqslant 1\}$i.i.d.,且

$$\mathsf{P}(X_n=1)=\mathsf{P}(X_n=-1)=\frac{1}{2}.$$

设 T 为$\{X_n,n\geqslant 1\}$的停时.令

$$Y_n=\begin{cases}X_n, & n\leqslant T,\\ -X_n, & n>T,\end{cases}\quad n\geqslant 1,$$

则$\{Y_n\}$与$\{X_n\}$同分布.

令 $S_n=X_1+\cdots+X_n$,$W_n=Y_1+\cdots+Y_n$,$n\geqslant 1$,则 $S=\{S_n,n\geqslant 1\}$与 $W=\{W_n,n\geqslant 1\}$的轨道到时刻 T 为止是相同的,在 T 之后,W 的轨道是 S 的轨道关于水平线 $y=S_T$ 的反射.本题的结论通常称为反射原理.

2-6. 设 $X=\{X_n,n\geqslant 1\}$为伯努利过程,每次试验成功的概率为 p,$N_n=X_1+\cdots+X_n$,$n\geqslant 1$,$N_0=0$,则对任意满足 $\sum_{n=0}^{\infty}|f(n)|<\infty$的数列$\{f(n),n\geqslant 1\}$有

$$\mathsf{E}\left[\sum_{n=0}^{\infty}f(N_n)\right]=\frac{1}{p}\sum_{n=0}^{\infty}f(n).$$

§1.3 泊松过程

定义 随机过程 $N=\{N(t),t\geqslant 0\}$,$N(0)=0$ 称为计数过程或点过程,若其样本函数以概率 1 为只取非负整数的右连续单调增加函数.对计数过程 N,若 $\Delta N(t)=N(t)-N(t^-)\leqslant 1$,则称 N 为简单的,这里 $N(t^-)$是 N 在 t 点的左极限.事实上,计数过程的样本函数是从零开始的跃度为正整数的阶梯函数,简单性意味着阶梯函数的跃度只是 1.

若用 $N(t)$表示某电话交换台在时间区间$(0,t]$中接到的电话呼唤的累计次数,则 $N=\{N(t),t\geqslant 0\}$就是计数过程,对电话呼唤次数进行累计的计数过程.这也就是计数过程的名称的由来.对 $0\leqslant s<t$,$N(t)-N(s)$就表示在$(s,t]$中发生的电话呼唤次数.$\Delta N(t)$表示这种随机事件在时刻 t 发生的呼唤次数.$\Delta N(t)=k$ 即表示在时刻 t 发生了 k 次呼唤.

计数的对象不仅仅是来到的电话呼唤,也可以是到某商店的顾客数,到达某机场降落的飞机数,某放射性物质在放射性蜕变中发射出的 α 粒子数,一次比赛中的进球数,某医院中出生的婴

儿数等等.总之,对某种随机事件的来到进行计数都可得到一个计数过程,而同一时刻只能发生一个来到的就是简单计数过程.

因此,我们称

$$T_k=\inf\{t\geqslant 0:\ N(t)\geqslant k\},\quad k\geqslant 0 \tag{1}$$

为第 k 次来到时刻.对简单计数过程则有

$$T_k=\inf\{t\geqslant 0:N(t)=k\},\quad k\geqslant 1, \tag{2}$$

这时 T_k 也称为第 k 个跳跃时刻.为方便起见,记 $T_0=0$.我们又称

$$W_k=T_k-T_{k-1},\quad k\geqslant 1 \tag{3}$$

为第 k 个等待时间,第 $k-1$ 次来到之后等第 k 次来到所需的等待时间.计数过程 N 也可以用来到时刻序列 $\{T_k,k\geqslant 1\}$ 构造:

$$N(t)=\sum_{k=1}^{\infty}1_{\{T_k\leqslant t\}},\quad t\geqslant 0 \tag{4}$$

或

$$N(t)=\begin{cases}0, & t<T_1,\\ k, & T_k\leqslant t<T_{k+1},k\geqslant 1,\end{cases}\quad t\geqslant 0. \tag{5}$$

由于 $T_k=W_1+\cdots+W_k,k\geqslant 1$,计数过程 N 也可以用等待时间序列 $\{W_k,k\geqslant 1\}$ 通过(5)构造.

定义 计数过程 $N=\{N(t),t\geqslant 0\}$ 称为时齐泊松过程,若它是平稳独立增量过程,且增量服从泊松分布:

$$\mathsf{P}(N(t)-N(s)=k)=\frac{[\lambda(t-s)]^k}{k!}\mathrm{e}^{-\lambda(t-s)},\quad 0\leqslant s<t,\quad \lambda>0. \tag{6}$$

λ 称为 N 的参数或强度.时齐泊松过程一般简称为泊松过程.

定理 3.1 泊松过程是一个简单计数过程.

证 当 $n\to\infty$ 时

$$\begin{aligned}&\mathsf{P}\{\text{存在 } s\in(0,t]\text{使 } \Delta N(s)\geqslant 2\}\\ &\leqslant\mathsf{P}\Big(\bigcup_{k=1}^{n}\Big\{N\Big(\frac{k}{n}t\Big)-N\Big(\frac{k-1}{n}t\Big)\geqslant 2\Big\}\Big)\\ &\leqslant\sum_{k=1}^{n}\mathsf{P}\Big(N\Big(\frac{k}{n}t\Big)-N\Big(\frac{k-1}{n}t\Big)\geqslant 2\Big)\\ &=\sum_{k=1}^{n}\Big(1-\mathrm{e}^{-\lambda t/n}-\frac{\lambda t}{n}\mathrm{e}^{-\lambda t/n}\Big)=O\Big(\frac{1}{n}\Big)\to 0.\end{aligned}$$

$\left(\text{这里 } O(x)\text{表示}\dfrac{O(x)}{x}\text{为有界量.}\right)$因此

$$\begin{aligned}&\mathsf{P}\{\text{对一切 } s\in(0,t],\Delta N(s)\leqslant 1\}=1,\\ &\mathsf{P}\{\text{对一切 } s>0,\Delta N(s)\leqslant 1\}=1.\quad\square\end{aligned}$$

定理 3.2 设 $N=\{N(t),t\geqslant 0\}$ 是不恒为零的简单计数过程，又是平稳独立增量过程，则 N 的增量必服从泊松分布，即 N 为泊松过程.

证 我们先证

$$\mathrm{P}(N(t)=0)=\mathrm{e}^{-\lambda t},\quad \lambda>0. \tag{7}$$

令 $f(t)=\mathrm{P}(N(t)=0),t\geqslant 0$. 由于 $\{N(t+s)=0\}=\{N(s)=0\}\cap\{N(t+s)-N(s)=0\}$, $s,t\geqslant 0$，故

$$\begin{aligned} f(t+s)&=\mathrm{P}(N(s)=0)\mathrm{P}(N(t+s)-N(s)=0) \qquad &(\text{增量独立})\\ &=\mathrm{P}(N(s)=0)\mathrm{P}(N(t)=0) &(\text{增量平稳})\\ &=f(s)f(t). \end{aligned} \tag{8}$$

又由 $f(t)$ 的定义，$0\leqslant f(t)\leqslant 1$，$f(t)$ 对 t 不增，故(8)的解为

$$f(t)\equiv 0 \quad 或 \quad f(t)=\mathrm{e}^{-\lambda t},\lambda\geqslant 0.$$

$f(t)\equiv 0$ 是不可能的，否则 $\mathrm{P}(N(t+s)-N(s)\geqslant 1)=1-f(t)=1$，从而对任意 $n\geqslant 1$ 取 $0<t_1<t_2<\cdots<t_{n-1}<t$ 并注意到

$$\begin{aligned} N(t)=(N(t)-N(t_{n-1}))+(N(t_{n-1})-N(t_{n-2}))+\cdots\\ +(N(t_2)-N(t_1))+N(t_1), \end{aligned}$$

得 $\mathrm{P}(N(t)\geqslant n)=1$. 由于 n 可以是任意的，所以可得 $\mathrm{P}(N(t)=+\infty)=1$. 这是不可能的. 所以只能是 $f(t)=\mathrm{e}^{-\lambda t},\lambda\geqslant 0$. 但 $\lambda=0$ 也是不可能的，因为 $f(t)\equiv 1$ 意味着 N 恒为零.

其次，我们证明

$$\lim_{t\to 0}\frac{1}{t}\mathrm{P}(N(t)\geqslant 2)=0. \tag{9}$$

((9)称为普通性.)记 $h(t)=\mathrm{P}(N(t)\geqslant 2),t\geqslant 0$，则 $h(t)$ 是 t 的不减函数，且

$$0\leqslant\frac{h(t)}{t}\leqslant h\left(\frac{1}{2^n}\right)\bigg/\frac{1}{2^{n+1}},\quad \frac{1}{2^{n+1}}<t<\frac{1}{2^n}.$$

所以为证明(9)，只需证明

$$\lim_{n\to\infty}2^n h\left(\frac{1}{2^n}\right)=0. \tag{10}$$

记 $A_n=\bigcup_{k=1}^{2^n}\left\{N\left(\frac{k}{2^n}\right)-N\left(\frac{k-1}{2^n}\right)\geqslant 2\right\}$，则 $A_{n+1}\subset A_n,n\geqslant 1$. 注意到 $\bigcap_{n=1}^{\infty}A_n$ 是不可能事件，即一切 $A_n,n\geqslant 1$ 不可能同时发生，因为 N 是简单的，只要 n 充分大，$N\left(\frac{k}{2^n}\right)-N\left(\frac{k-1}{2^n}\right)$ 只可能取 0 或 1 两个值. 所以，$\mathrm{P}(A_n)\to 0$. 另一方面，由平稳独立增量性

$$1-\mathrm{P}(A_n)=\mathrm{P}\left(\bigcap_{k=1}^{2^n}\left\{N\left(\frac{k}{2^n}\right)-N\left(\frac{k-1}{2^n}\right)<2\right\}\right)$$

$$= \prod_{k=1}^{2^n} \mathrm{P}\left(N\left(\frac{k}{2^n}\right) - N\left(\frac{k-1}{2^n}\right) < 2\right) = \left[1 - h\left(\frac{1}{2^n}\right)\right]^{2^n},$$

注意到 $\log(1+\delta) \leqslant \delta, \delta > -1$，及对充分小的 t 有 $h(t) < 1$，

$$0 \geqslant -2^n h\left(\frac{1}{2^n}\right) \geqslant 2^n \log\left[1 - h\left(\frac{1}{2^n}\right)\right] = \log[1 - \mathrm{P}(A_n)],$$

从而(10)成立. 因为

$$\mathrm{P}(N(t)=1) = 1 - \mathrm{P}(N(t)=0) - \mathrm{P}(N(t) \geqslant 2) = 1 - \mathrm{e}^{-\lambda t} - h(t),$$

故由(9)得

$$\lim_{t \to \infty} \frac{1}{t} \mathrm{P}(N(t)=1) = \lambda. \tag{11}$$

记 $G(t,z) = \mathrm{E}(z^{N(t)}) = \sum_{n=0}^{\infty} \mathrm{P}(N(t)=n) z^n, |z| < 1$，则

$$\begin{aligned} G(t+s,z) &= \mathrm{E}(z^{N(t+s)}) = \mathrm{E}(z^{N(t)} z^{N(t+s)-N(t)}) \\ &= \mathrm{E}(z^{N(t)}) \mathrm{E}(z^{N(t+s)-N(t)}) = G(t,z) G(s,z). \end{aligned}$$

又因 $0 < G(t,z) \leqslant 1$，$G(t,z)$ 对 t 是不增的，所以 $G(t,z) = \mathrm{e}^{tg(z)}$. 又由(8)，(9)及(11)

$$\begin{aligned} g(z) &= \lim_{t \downarrow 0} \frac{G(t,z) - 1}{t} \\ &= \lim_{t \downarrow 0} \left\{ \frac{1}{t} [\mathrm{P}(N(t)=0) - 1] + \frac{z}{t} \mathrm{P}(N(t)=1) + \frac{1}{t} \sum_{k=2}^{\infty} \mathrm{P}(N(t)=k) z^k \right\} \\ &= -\lambda + \lambda z = \lambda(z-1). \end{aligned}$$

所以

$$G(t,z) = \mathrm{e}^{t\lambda(z-1)} = \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} \mathrm{e}^{-\lambda t} z^n,$$

$$\mathrm{P}(N(t)=n) = \frac{(\lambda t)^n}{n!} \mathrm{e}^{-\lambda t}.$$

即 N 的增量的分布为泊松分布，故 N 为泊松过程. □

在定理 3.2 的证明中，我们实际上利用了概率分布列的母函数. 设 $\{p_n, n \geqslant 0\}$ 为一概率分布列，称

$$G(z) = \sum_{n=0}^{\infty} p_n z^n, \quad |z| \leqslant 1$$

为它的母函数. 显然，概率分布列与其母函数是一一对应的. 若随机变量 X 为整值随机变量（即 X 只取非负整数值），以 $\{p_n, n \geqslant 0\}$ 为其分布列：$\mathrm{P}(X=n) = p_n$，则

$$G(z) = \mathrm{E}[z^X]$$

也称为 X 的母函数. 不难算得

$$\mathrm{E}\,X=G'(1),\quad \mathrm{E}\,X^2=G''(1)+G'(1).$$

在定理 3.2 的证明中,我们也可看到,泊松过程的强度 λ 有多种含义:

1) 单位时间的平均来到数

$$\frac{1}{t}\mathrm{E}[N(t)]=\mathrm{E}[N(1)]=\lambda;$$

2) 样本函数的平均来到数(由强大数定律得)

$$\lim_{t\to\infty}\frac{1}{t}N(t)=\lambda;$$

3) 来到一个(或至少一个)的速率(由(9)及(11)得)

$$\begin{aligned}&\lim_{h\to0^+}\frac{1}{h}\mathrm{P}(N(t+h)-N(t)=1)\\&=\lim_{h\to0^+}\frac{1}{h}\mathrm{P}(N(t+h)-N(t)\geqslant1)=\lambda.\end{aligned}$$

定理 3.3 设 $N=\{N(t),t\geqslant0\}$ 为泊松过程,参数为 λ,则等待时间序列 $\{W_j,j\geqslant1\}$ 为i.i.d.随机变量序列且同指数分布

$$\mathrm{P}(W_j>x)=\mathrm{e}^{-\lambda x},\quad x>0.\tag{12}$$

来到时刻 $T_k,k\geqslant1$,服从 $\Gamma(k,\lambda)$ 分布:

$$\mathrm{P}(T_k\leqslant t)=\int_0^t\frac{(\lambda s)^{k-1}}{(k-1)!}\lambda\mathrm{e}^{-\lambda s}\mathrm{d}s.\tag{13}$$

证 任取 $k\geqslant1$. 由泊松过程的简单性可知,$(T_1,\cdots,T_k)$ 的取值范围可定为 $\{(t_1,\cdots,t_k):0<t_1<t_2<\cdots<t_k\}$. 当 $0<t_1<\cdots<t_k$ 时,

$$\begin{aligned}&\mathrm{P}(T_j>t_j,1\leqslant j\leqslant k)=\mathrm{P}(N(t_j)\leqslant j-1,1\leqslant j\leqslant k)\\&=\sum_{\substack{i_1+\cdots+i_j\leqslant j-1\\1\leqslant j\leqslant k}}\mathrm{P}(N(t_1)=i_1,N(t_2)-N(t_1)=i_2,\cdots,N(t_k)-N(t_{k-1})=i_k)\\&=\sum_{\substack{i_1+\cdots+i_j\leqslant j-1\\1\leqslant j\leqslant k}}\frac{(\lambda t_1)^{i_1}}{i_1!}\cdot\frac{[\lambda(t_2-t_1)]^{i_2}}{i_2!}\cdots\frac{[\lambda(t_k-t_{k-1})]^{i_k}}{i_k!}\mathrm{e}^{-\lambda t_k}.\end{aligned}\tag{14}$$

在上式中求偏导数 $\frac{\partial^k}{\partial t_1\cdots\partial t_k}$,左边得 $(T_1,\cdots,T_k)$ 的联合分布密度乘以 $(-1)^k$. 考察右边的情形,(14)右边的和式中 $i_1=0$,而 i_2 可取 0 或 1. 对应 $i_2=0$ 的项,求 $\frac{\partial}{\partial t_1}$ 后即为 0,故只需考虑 $i_1=0$, $i_2=1$ 的项. 此时 i_3 可取为 0 或 1,但对应 $i_3=0$ 的项,求 $\frac{\partial^2}{\partial t_1\partial t_2}$ 后即为零,故只保留 $i_1=0,i_2=1,i_3=1$ 的项. 依次类推可知只有 $i_1=0,i_2=1,\cdots,i_k=1$ 的项求 $\frac{\partial^k}{\partial t_1\cdots\partial t_k}$ 后不为零,且求导结果是 $(-1)^k\lambda^k\mathrm{e}^{-\lambda t_k}$. 由此 $(T_1,\cdots,T_k)$ 的联合分布密度为

$$f_{T_1,\cdots,T_k}(t_1,\cdots,t_k)=\begin{cases}\lambda^k e^{-\lambda t_k}, & 0<t_1<\cdots<t_k,\\ 0, & \text{其它}.\end{cases}$$

由于 $T_j=\sum_{i=1}^{j} W_i, j\geqslant 1$，因而 $(W_1,\cdots,W_k)$ 的联合分布密度为

$$f_{W_1,\cdots W_k}(w_1,\cdots,w_k)=\lambda^k e^{-\lambda\sum_{j=1}^{k}w_j}=\prod_{j=1}^{k}\lambda e^{-\lambda w_j},\ w_1,\cdots,w_k>0.$$

所以 $W_1,\cdots,W_k$ 为 i.i.d. 的，且以(12)为其一维分布.

利用 $T_k=\inf\{t:N(t)=k\}$，$\{T_k\leqslant t\}=\{N(t)\geqslant k\}$ 可得

$$\mathsf{P}(T_k\leqslant t)=\mathsf{P}(N(t)\geqslant k)=\sum_{j=k}^{\infty}\frac{(\lambda t)^j}{j!}e^{-\lambda t},$$

$$\frac{\partial \mathsf{P}(T_k\leqslant t)}{\partial t}=\sum_{j=k}^{\infty}\left[\lambda\frac{(\lambda t)^{j-1}}{(j-1)!}e^{-\lambda t}-\lambda\frac{(\lambda t)^j}{j!}e^{-\lambda t}\right]=\lambda\frac{(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t},$$

故(13)成立. □

定理 3.3 给出了泊松过程的一种构造法：从一列独立同指数分布的随机变量 $\{W_k,k\geqslant 1\}$（作为等待时间序列）出发，令 $T_k=W_1+\cdots+W_k, k\geqslant 1$，再通过(5)定义出泊松过程 N. 我们知道，指数分布具有无后效性：对任意的 $s,t\geqslant 0$

$$\mathsf{P}(W_k>t+s\mid W_k>t)=\mathsf{P}(W_k>s). \tag{15}$$

这无后效性与泊松过程的平稳独立增量性相联系. 事实上不难证明，任一只取正值的连续型随机变量若有无后效性(15)，其分布必为指数分布.

定理 3.4 设 $N=\{N(t),t\geqslant 0\}$ 是参数为 λ 的泊松过程，$T_k, k\geqslant 1$ 为其来到时刻，则对任意的 $[0,\infty)$ 上的可积函数 f 有

$$\mathsf{E}\left[\sum_{n=1}^{\infty}f(T_n)\right]=\lambda\int_0^{\infty}f(t)\mathrm{d}t. \tag{16}$$

证 先设 f 非负. 由(13)

$$\mathsf{E}[f(T_k)]=\lambda\int_0^{\infty}f(t)\frac{(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t}\mathrm{d}t,$$

$$\sum_{k=1}^{\infty}\mathsf{E}[f(T_k)]=\lambda\int_0^{\infty}f(t)\sum_{k=1}^{\infty}\frac{(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t}\mathrm{d}t=\lambda\int_0^{\infty}f(t)\mathrm{d}t.$$

对一般的 f，将已证结果用于 $f^+=\max(f,0)$ 及 $f^-=\max(-f,0)$，即可知(16)对 $f=f^+-f^-$ 成立. □

定义 设 $X_1,\cdots,X_n$ 为取自分布 F 的一个样本，即 $X_1,\cdots,X_n$ 为 i.i.d. 随机变量且都具有分布 F. 又 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 为 $X_1,\cdots,X_n$ 按取值大小的一个排列，称 $(X_{(1)},X_{(2)},\cdots,X_{(n)})$ 为 $(X_1,\cdots,X_n)$ 的次序统计量.

当 $F(x)$ 以 $f(x)$ 为密度时，$(X_{(1)},\cdots,X_{(n)})$ 的分布密度为

$$p(x_1,\cdots,x_n)=\begin{cases}n!\ \prod_{j=1}^{n}f(x_j), & x_1<\cdots<x_n,\\ 0, & \text{其它}.\end{cases}$$

特别若 F 为$(0,t]$上的均匀分布,则次序统计量的分布密度为

$$p(x_1,\cdots,x_n)=\begin{cases}n!\ t^{-n}, & 0<x_1<\cdots<x_n<t,\\ 0, & \text{其它}.\end{cases}$$

定理 3.5 设 $N=\{N(t),t\geqslant 0\}$是强度为 λ 的泊松过程,$\{T_n,n\geqslant 1\}$为其跳跃时刻序列,则

$$\begin{aligned}&\mathrm{P}(T_j\in(s_j,t_j],1\leqslant j\leqslant n\mid N(t)=n)\\&=\frac{n!}{t^n}\prod_{j=1}^{n}(t_j-s_j),\ 0<s_1<t_1<\cdots<s_n<t_n<t,\end{aligned}\tag{17}$$

即已知$(0,t]$中 N 发生 n 次跳跃的条件下,n 个跳跃时刻$(T_1,\cdots,T_n)$与 n 个独立的$(0,t]$上均匀分布的随机变量的次序统计量有相同分布.

证 当 $0=t_0<s_1<t_1<\cdots<s_n<t_n<t$ 时,

$$\begin{aligned}&\mathrm{P}(T_j\in(s_j,t_j],1\leqslant j\leqslant n\mid N(t)=n)\\&=\mathrm{P}(N(t_j)-N(s_j)=1,N(s_j)-N(t_{j-1})=0,\\&\qquad 1\leqslant j\leqslant n,N(t)-N(t_n)=0\mid N(t)=n)\\&=\prod_{j=1}^{n}[\lambda(t_j-s_j)]\mathrm{e}^{-\lambda(t_j-s_j)}\mathrm{e}^{-\lambda(s_j-t_{j-1})}\mathrm{e}^{-\lambda(t-t_n)}\Big/\left(\frac{(\lambda t)^n}{n!}\mathrm{e}^{-\lambda t}\right)\\&=\frac{n!}{t^n}\prod_{j=1}^{n}(t_j-s_j).\quad\square\end{aligned}$$

直观地说,对泊松过程 N,若已知 $N(t)=n$,那么可以认为这 n 个跳跃时刻在$(0,t]$上的分布就如同$(0,t]$上均匀分布的容量为 n 的样本一样.这一性质通常也称为泊松过程的完全随机性.

例 3.1 设到达某测试仪表的脉冲数是强度为 λ 的泊松过程 $N=\{N(t),t\geqslant 0\}$,脉冲的电压峰值为 i.i.d. 随机变量列$\{D_i,i\geqslant 1\}$,且与 $N=\{N(t),t\geqslant 0\}$独立.脉冲到达后随时间以指数衰减,即在到达后经过时间 s 脉冲衰减为$D_i\mathrm{e}^{-\alpha s}$.试求在时刻 t 仪表能测量到的来到脉冲电压总和的均值.

解 在时刻 t 测到的脉冲电压总和为

$$D(t)=\sum_{i=1}^{N(t)}D_i\mathrm{e}^{-\alpha(t-T_i)}.$$

$D(t)$的均值为

$$\begin{aligned}\mathrm{E}\,D(t)&=\mathrm{E}\left[\sum_{i=1}^{N(t)}D_i\mathrm{e}^{-\alpha(t-T_i)}\right]\\&=\sum_{j=1}^{\infty}\mathrm{E}\left[\sum_{i=1}^{j}D_i\mathrm{e}^{-\alpha(t-T_i)}\mid N(t)=j\right]\mathrm{P}(N(t)=j)\\&=\mathrm{e}^{-\alpha t}\sum_{j=1}^{\infty}\mathrm{E}\left[\sum_{i=1}^{j}D_i\mathrm{e}^{-\alpha T_i}\mid N(t)=j\right]\mathrm{P}(N(t)=j)\\&=\mathrm{e}^{-\alpha t}\,\mathrm{E}[D_1]\sum_{j=1}^{\infty}\mathrm{E}\left[\sum_{i=1}^{j}\mathrm{e}^{\alpha U_i}\right]\mathrm{P}(N(t)=j)\end{aligned}$$

$$=\mathrm{e}^{-\alpha t}\,\mathrm{E}[D_1]\sum_{j=1}^{\infty}j\,\frac{1}{t}\int_0^t \mathrm{e}^{\alpha u}\mathrm{d}u\,\mathrm{P}(N(t)=j)$$

$$=\mathrm{e}^{-\alpha t}\,\mathrm{E}[D_1]\frac{\mathrm{e}^{\alpha t}-1}{\alpha t}\lambda t=\frac{\lambda\ \mathrm{E}[D_1]}{\alpha}(1-\mathrm{e}^{-\alpha t}),$$

其中$\{U_i,i\geqslant 1\}$为 i.i.d.$(0,t]$上均匀分布随机变量序列.这里我们利用了定理 3.5.

下面我们利用定理 3.4,提供另一种计算$\mathrm{E}[D(t)]$的方法.若取

$$f(s)=1_{[0,t]}(s)\mathrm{e}^{-\alpha(t-s)},$$

则

$$D(t)=\sum_{i=1}^{N(t)}D_i\mathrm{e}^{-\alpha(t-T_i)}=\sum_{i=1}^{\infty}1_{\{T_i\leqslant t\}}D_i\mathrm{e}^{-\alpha(t-T_i)}=\sum_{i=1}^{\infty}D_if(T_i),$$

$$\mathrm{E}[D(t)\}=\sum_{i=1}^{\infty}\mathrm{E}[D_if(T_i)]=\mathrm{E}[D_1]\mathrm{E}\Big[\sum_{i=1}^{\infty}f(T_i)\Big]$$

$$=\mathrm{E}[D_1]\lambda\int_0^{\infty}f(s)\mathrm{d}s=\mathrm{E}[D_1]\lambda\int_0^t \mathrm{e}^{-\alpha(t-s)}\mathrm{d}s$$

$$=\frac{\lambda}{\alpha}(1-\mathrm{e}^{-\alpha t})\mathrm{E}[D_1].\quad\square$$

例 3.2 设乘客按强度为 λ 的泊松过程来到公共汽车站候车.每隔时间 T 开出一辆公共汽车,将站上候车的乘客全部载走.现在为了缩短乘客的候车时间,在原来两班车之间再增开一辆车,即在区间$(0,T)$中选取一个时刻 t,在时刻 t 增开一辆车将$(0,t]$中来到的乘客载走,而在时刻 T 开出的班车只需将$(t,T]$中来到的乘客载走.应该如何选取 t,使得$(0,T]$中来到的全部乘客的总的平均等待时间最短.

解 以$\{N(t)\}$记顾客来到的泊松过程.先计算$(0,t]$中来到的乘客的等待时间总和的平均值(乘客等到时刻 t 为止).设 $N(t)=n$,可以认为这 n 个乘客来到车站的时刻服从$(0,t)$上的均匀分布,因此每个乘客的平均候车时间为 $t/2$,这 n 个乘客的总的平均候车时间为 $nt/2$.这样,$(0,t]$中来到的乘客的总的平均候车时间

$$\sum_{n=1}^{\infty}\frac{nt}{2}\mathrm{P}(N(t)=n)=\sum_{n=1}^{\infty}\frac{nt}{2}\frac{(\lambda t)^n}{n!}\mathrm{e}^{-\lambda t}=\frac{\lambda t^2}{2}.$$

由泊松过程的平稳增量性,在$(t,T]$中来到的乘客的总的平均候车时间为 $\lambda(T-t)^2/2$.因此,我们只要求 t,使得

$$\frac{\lambda t^2}{2}+\frac{\lambda(T-t)^2}{2}$$

达到最小.容易求得 $t=T/2$. $\square$

定理 3.6 设 $N=\{N(t),t\geqslant 0\}$是强度为 λ 的泊松过程,$\{T_n,n\geqslant 1\}$为其来到时刻序列.对任一 $t>0$,令

$$R(t)=T_{N(t)+1}-t,\qquad A(t)=t-T_{N(t)},\tag{18}$$

则 $R(t)$与 $A(t)$相互独立,且

$$\mathsf{P}(R(t)\leqslant x)=1-\mathrm{e}^{-\lambda x},\qquad x\geqslant 0,\tag{19}$$

$$\mathsf{P}(A(t)\leqslant x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & 0\leqslant x<t,\\ 1, & x\geqslant t.\end{cases}\tag{20}$$

证 从 $A(t)$的定义容易看出 $A(t)\leqslant t$,所以当 $y\leqslant t$ 时

$$\{R(t)>x,A(t)\geqslant y\}=\{在(t-y,t+x]中\ N\ 无跳跃\},$$

$$\mathsf{P}(R(t)>x,A(t)\geqslant y)=\mathsf{P}(N(t+x)-N(t-y)=0)=\mathrm{e}^{-\lambda(x+y)}.$$

因此

$$\mathsf{P}(R(t)>x,A(t)\geqslant y)=\begin{cases}\mathrm{e}^{-\lambda(x+y)}, & x\geqslant 0,0\leqslant y\leqslant t,\\ 0, & y>t.\end{cases}\tag{21}$$

特别地,在(21)中令 $y=0$ 得(19),令 $x=0$ 得(20). 再由(21)可知,对 $x\geqslant 0,y\geqslant 0$,

$$\mathsf{P}(R(t)>x,A(t)\geqslant y)=\mathrm{e}^{-\lambda(x+y)}1_{[0,t]}(y)=\mathsf{P}(R(t)>x)\mathsf{P}(A(t)\geqslant y).$$

所以 $R(t)$与 $A(t)$是独立的. □

设以泊松过程 N 表示来到某车站的公共汽车总数. 在时刻 t 乘客来到车站,则在 t 之前最近的一辆公共汽车是在时刻 $T_{N(t)}$开走的,已开走了时间 $A(t)$,而下一辆公共汽车将在时刻 $T_{N(t)+1}$到来,乘客将等待时间 $R(t)$,而这两辆车间隔时间 $\beta(t)=R(t)+A(t)$,其平均值

$$\begin{aligned}\mathsf{E}[\beta(t)]&=\mathsf{E}[R(t)]+\mathsf{E}[A(t)]\\&=\frac{1}{\lambda}+\int_0^t\lambda x\mathrm{e}^{-\lambda x}\mathrm{d}x=\frac{1}{\lambda}+\frac{1}{\lambda}(1-\mathrm{e}^{-\lambda t}).\end{aligned}\tag{22}$$

另一方面,按泊松过程的性质,其相邻两次来到间隔的均值为 $\mathsf{E}\,T_1=1/\lambda$. 但(22)表明 $\mathsf{E}[\beta(t)]>1/\lambda$,甚至

$$\lim_{t\to\infty}\mathsf{E}[\beta(t)]=\frac{2}{\lambda}=2\ \mathsf{E}\,T_1.$$

这一事实,表面看来是一个矛盾,实际上并非如此. 尽管泊松过程的来到间隔是一个随机变量,其均值应为 $1/\lambda$,但在固定的一个时刻 t 来观察包含 t 的那个来到间隔,它的分布已不同于原有的来到间隔,原有的长长短短的来到间隔排在时间轴上,一般说来,长的间隔有更多的可能包含 t,短的间隔包含 t 的可能较小,所以跨 t 的那个间隔更多的是偏向长的间隔,它的均值 $\mathsf{E}[\beta(t)]$ 也就大于原有的均值 $1/\lambda$. 这也就是人们在等车时总感到间隔比较长的一个原因. 值得注意的是,不论乘客在哪一个时刻 t 到车站,他等待下一辆公共汽车来到的时间 $R(t)$同样服从参数为 λ 的指数分布.

定理 3.7 设 $N=\{N(t),t\geqslant 0\}$是参数为 λ 的泊松过程,$Y=\{Y_i,i\geqslant 1\}$为 i.i.d. 整值随机变量序列,

$$\mathsf{P}(Y_i=k)=a_k,\qquad k\geqslant 0,$$

且 Y 与 N 独立. 令

$$X(t)=\sum_{i=1}^{N(t)} Y_i, \quad t\geqslant 0, \tag{23}$$

则 X 为平稳独立增量过程,且其增量的母函数为

$$\mathrm{E}\{z^{(X(t)-X(s))}\}=\exp\{\lambda(t-s)(A(z)-1)\}, \quad 0\leqslant s<t, \tag{24}$$

其中 $A(z)$ 为 Y_1 的母函数:$A(z)=\sum_{k=0}^{\infty} a_k z^k$.

证 对 $0=t_0<t_1<\cdots<t_k, m_1,\cdots,m_k\in\mathbb{N}$,

$$\begin{aligned}
&\mathrm{P}(X(t_j)-X(t_{j-1})=m_j, 1\leqslant j\leqslant k)\\
&=\sum_{n_1,\cdots,n_k\geqslant 0}\mathrm{P}\Big(\bigcap_{j=1}^{k}\{X(t_j)-X(t_{j-1})=m_j\}\Big|\bigcap_{j=1}^{k}\{N(t_j)-N(t_{j-1})=n_j\}\Big)\\
&\quad\times\prod_{j=1}^{k}\mathrm{P}(N(t_j)-N(t_{j-1})=n_j)\\
&=\sum_{n_1,\cdots,n_k\geqslant 0}\mathrm{P}\Big(\sum_{i=n_1+\cdots+n_{j-1}+1}^{n_1+\cdots+n_j} Y_i=m_j, 1\leqslant j\leqslant k\Big)\prod_{j=1}^{k}\mathrm{P}(N(t_j)-N(t_{j-1})=n_j)\\
&=\sum_{n_1,\cdots,n_k\geqslant 0}\prod_{j=1}^{k}\Big[P\Big(\sum_{i=n_1+\cdots+n_{j-1}+1}^{n_1+\cdots+n_j} Y_i=m_j\Big)\mathrm{P}(N(t_j)-N(t_{j-1})=n_j)\Big]\\
&=\sum_{n_1,\cdots,n_k\geqslant 0}\prod_{j=1}^{k}\Big[\mathrm{P}\Big(\sum_{i=1}^{n_j} Y_i=m_j\Big)\mathrm{P}(N(t_j)-N(t_{j-1})=n_j)\Big]\\
&=\prod_{j=1}^{k}\sum_{n_j=0}^{\infty}\Big[\mathrm{P}\Big(\sum_{i=1}^{n_j} Y_i=m_j\Big)\mathrm{P}(N(t_j)-N(t_{j-1})=n_j)\Big]\\
&=\prod_{j=1}^{k}\mathrm{P}(X(t_j)-X(t_{j-1})=m_j).
\end{aligned}$$

故 X 有平稳独立增量,且

$$\begin{aligned}
\mathrm{E}(z^{X(t)})&=\sum_{k=0}^{\infty}\mathrm{E}(z^{X(t)}\mid N(t)=k)\mathrm{P}(N(t)=k)\\
&=\sum_{k=0}^{\infty}\mathrm{E}(z^{\sum_{j=1}^{k}Y_j})\mathrm{P}(N(t)=k)\\
&=\sum_{k=0}^{\infty}\Big(\prod_{j=1}^{k}\mathrm{E}(z^{Y_j})\Big)\mathrm{P}(N(t)=k)\\
&=\sum_{k=0}^{\infty}[A(z)]^k\frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t}=\exp(\lambda t(A(z)-1)).
\end{aligned}$$

因此(24)成立.这里,我们用到了下列事实:独立随机变量和的母函数等于各变量母函数的积.
□

(23)定义的过程 X 称为复合泊松过程.一般,其中的 i.i.d.序列 $\{Y_i, i\geqslant 1\}$ 可以是取任意值的随机变量序列.它在经典的保险理论中有重要的应用.例如,假设发生火灾的累计次数为泊松过程 $\{N(t)\}$,第 i 次火灾后支付的赔偿金为 Y_i,则到时刻 t 累计的赔偿金总数即为

$$X(t)=\sum_{i=1}^{N(t)} Y_i.$$

一般还假设 $Y=\{Y_i,i\geqslant 1\}$ i.i.d.,且与 N 相互独立.这样,$X=\{X(t),t\geqslant 0\}$ 就是复合泊松过程了.

定理 3.8 设 $N=\{N(t),t\geqslant 0\}$ 是参数为 λ 的泊松过程,$Y=\{Y_i,i\geqslant 1\}$ 为 i.i.d. 随机变量序列,

$$\mathrm{P}(Y_i=1)=p, \qquad \mathrm{P}(Y_i=0)=1-p, \tag{25}$$

且 Y 与 N 独立.令

$$N_1(t)=\sum_{i=1}^{N(t)} Y_i, \quad N_2(t)=\sum_{i=1}^{N(t)}(1-Y_i), \quad t\geqslant 0, \tag{26}$$

则 $N_1=\{N_1(t),t\geqslant 0\}$ 与 $N_2=\{N_2(t),t\geqslant 0\}$ 为相互独立的泊松过程,参数分别为 λp 与 $\lambda(1-p)$.

证 用定理 3.7 即知,N_1 与 N_2 都是平稳独立增量过程.由假设(25),用(24)可知,$N_1(t)$ 与 $N_2(t)$ 的母函数分别为 $\exp\{\lambda pt(z-1)\}$ 与 $\exp\{\lambda(1-p)(z-1)\}$.因此,$N_1$ 与 N_2 分别是参数为 λp 与 $\lambda(1-p)$ 的泊松过程.再证它们相互独立.

对任意 $k\geqslant 1, 0\leqslant s_1<t_s<s_2<t_2<\cdots<s_k<t_k, m_1,n_1,\cdots,m_k,n_k\in\mathbb{N}$

$$\begin{aligned}
&\mathrm{P}(N_1(t_j)-N_1(s_j)=m_j, N_2(t_j)-N_2(s_j)=n_j, 1\leqslant j\leqslant k)\\
&=\mathrm{P}(N(t_j)-N(s_j)=m_j+n_j, 1\leqslant j\leqslant k)\mathrm{P}(N_1(t_j)-N_1(s_j)=m_j,\\
&\qquad N_2(t_j)-N_2(s_j)=n_j, 1\leqslant j\leqslant k \mid N(t_j)-N(s_j)=m_j+n_j, 1\leqslant j\leqslant k)\\
&=\prod_{j=1}^{k}\left\{\frac{[\lambda(t_j-s_j)]^{m_j+n_j}}{(m_j+n_j)!}\mathrm{e}^{-\lambda(t_j-s_j)}\right\}\prod_{j=1}^{k}\left\{\mathrm{C}_{m_j+n_j}^{m_j}p^{m_j}(1-p)^{n_j}\right\}\\
&=\prod_{j=1}^{k}\left\{\frac{[\lambda p(t_j-s_j)]^{m_j}}{m_j!}\mathrm{e}^{-\lambda p(t_j-s_j)}\right\}\prod_{j=1}^{k}\left\{\frac{[\lambda(1-p)(t_j-s_j)]^{n_j}}{n_j!}\mathrm{e}^{-\lambda(1-p)(t_j-s_j)}\right\}\\
&=\mathrm{P}(N_1(t_j)-N_1(s_j)=m_j, 1\leqslant j\leqslant k)\mathrm{P}(N_2(t_j)-N_2(s_j)=n_j, 1\leqslant j\leqslant k).
\end{aligned}$$

这即表明 N_1 与 N_2 相互独立. □

定理 3.8 中的过程 N_1 或 N_2 称为 N 的稀疏过程,因为它们是将 N 的来到随机地剔除一部分,经过稀疏而得到的.计数过程的稀疏是实际中常会遇到的现象.例如,来到某商店门口的顾客数为一计数过程,到店门口后有的顾客进门,有的顾客不进门.进门的顾客数构成的过程就是前一计数过程的稀疏过程.再如,通过某十字路口的车辆数为一计数过程,过十字路口的车辆有的转弯,有的不转弯.转弯的车辆数构成的过程也是稀疏过程.

习 题

3-1. 设 $N=\{N_1(t),t\geqslant 0\}$ 与 $N_2=\{N_2(t),t\geqslant 0\}$ 为两个相互独立的泊松过程,参数分别为 λ_1 与 λ_2.令

$$N(t)=N_1(t)+N_2(t), \qquad t\geqslant 0.$$

证明：$N=\{N(t),t\geqslant 0\}$ 为泊松过程，参数为 $\lambda=\lambda_1+\lambda_2$.

3－2. 对泊松过程 $\{N(t),t\geqslant 0\}$，证明：若 $0<s<t$，$n\geqslant 1$，则

$$\mathrm{P}(N(s)=k\mid N(t)=n)=\mathrm{C}_n^k\left(\frac{s}{t}\right)^k\left(1-\frac{s}{t}\right)^{n-k},\ k=0,1,\cdots,n.$$

3－3. 设 $\{T_k,k\geqslant 1\}$ 是强度为 λ 的泊松过程 N 的来到时刻序列，f 为 $[0,\infty)$ 上的非负可积函数. 证明：

$$\mathrm{E}\left\{\exp\left[-\sum_{n=1}^{\infty}f(T_n)\right]\right\}=\exp\left\{-\lambda\int_0^{\infty}\left[1-\mathrm{e}^{-f(t)}\right]\mathrm{d}t\right\}.$$

并由此计算 $\mathrm{E}\left[-\sum_{n=1}^{\infty}f(T_n)\right]$ 及 $\mathrm{Var}\left[\sum_{n=1}^{\infty}f(T_n)\right]$.

3－4. 设 $N=\{N(t),t\geqslant 0\}$ 为计数过程，且对任意 $n(\geqslant 1)$ 个互不相交的区间 $(s_j,t_j]$，$j=1,\cdots,n$，$\sum_{j=1}^{n}(N(t_j)-N(s_j))$ 服从参数为 $\lambda\sum_{j=1}^{n}(t_j-s_j)$ 的泊松分布，$\lambda>0$，则 N 为泊松过程.

3－5. 设 $N=\{N(t),t\geqslant 0\}$ 是强度为 λ 的泊松过程，随机变量 T 与 N 独立，$\mathrm{P}(T>t)=\mathrm{e}^{-\mu t}$，$t>0$. 证明：

$$\mathrm{P}(N(T)=k)=\frac{\mu}{\lambda+\mu}\left(\frac{\lambda}{\lambda+\mu}\right)^k,\quad k=0,1,2,\cdots.$$

3－6. 某种核粒子按参数为 λ 的泊松过程来到一个计数器. 每个到达的粒子都使计数器关闭一段时间 r. 当一个粒子到达时，计数器如不处于关闭状态，它就被记录下来. 求在时间区间 $(t,t+r]$ $(t\geqslant r)$ 中记录下一个粒子的概率.

3－7. 某台仪器有 A,B 两个系统构成. 可能发生的故障有三种类型，发生第 i 类 $(i=1,2,3)$ 故障的累计次数构成三个相互独立的强度分别为 λ_i 的泊松过程. 若发生第一类故障，系统 A 无法正常运行；若发生第二类故障，系统 B 无法正常运行；若发生第三类故障，系统 A,B 都无法正常运行. 以 X_1 和 X_2 分别记 A,B 两个系统的寿命. 证明：

(1) X_1 和 X_2 的联合分布为

$$\mathrm{P}(X_1>s,X_2>t)=\exp\{-\lambda_1 s-\lambda_2 t-\lambda_3\max(s,t)\}$$

（这分布通常称为二元指数分布）.

(2) X_1 和 X_2 均服从指数分布.

3－8. 设 $N^{(k)}=\{N^{(k)}(t),t\geqslant 0\}$，$k\geqslant 1$ 为一列相互独立的泊松过程，$N^{(k)}$ 的强度为 λ_k，又 $\lambda=\sum_{k=1}^{\infty}\lambda_k<\infty$. 令

$$N(t)=\sum_{k=1}^{\infty}kN^{(k)}(t),\qquad t\geqslant 0.$$

证明：$N=\{N(t),t\geqslant 0\}$ 为复合泊松过程.

3－9. 考虑如图 1 所示的交通网络. 流入的是按所示强度的泊松过程，而在交会处车辆按图示的概率选择行走的方向. 描述三个出口处的交通流情况.

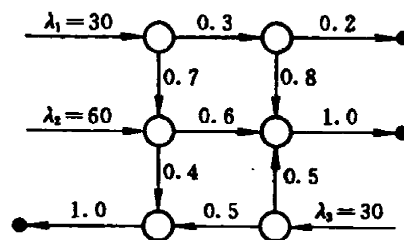


图 1

第二章　离散时间的马尔可夫链

§2.1　马尔可夫性

定义　设$\{X_n, n\geqslant 0\}$为一列只取非负整数值的随机变量.若对任意的$k\geqslant 1, 0\leqslant t_0<t_1<\cdots<t_{k+1}$及非负整数$i_0, i_1, \cdots, i_{k+1}$,

$$\mathrm{P}(X_{t_{k+1}}=i_{k+1}\mid X_{t_0}=i_0, X_{t_1}=i_1, \cdots, X_{t_k}=i_k)=\mathrm{P}(X_{t_{k+1}}=i_{k+1}\mid X_{t_k}=i_k) \tag{1}$$

(等式成立的意义是只要等式两边的条件概率有意义,等式就成立.今后凡是涉及条件概率的等式,其成立的意义都按这样理解,不再重复指明),则称$\{X_n, n\geqslant 0\}$为*离散时间离散状态的马尔可夫过程*,或*离散时间的马尔可夫链*.

$\{X_n, n\geqslant 0\}$描述了一个随时间而变化的随机系统的状态,事件$\{X_n=i\}$表示在时刻n系统处于状态i.离散状态实际上是指状态空间E为可列集,即全体状态可用非负整数进行编号.为方便起见,今后在一般的讨论中,除非另有特别声明,我们约定用非负整数集$\mathbb{N}=\{0,1,2,\cdots\}$作为状态空间,即$E=\mathbb{N}$(这些整数可看作它所代表的状态的编号),这也就是我们一开始假定随机变量只取非负整数值的原因.在实际应用中自然可以不受这个限制.例如,状态空间取为整数集$\mathbb{Z}=\{\cdots,-1,0,1,\cdots\}$也是常见的.离散状态的马尔可夫过程一般称为*马尔可夫链*.状态空间也可以是有限集,这时马尔可夫链就称为*有限状态的*,或称为*有限马尔可夫链*.一般,有限状态空间用$\{0,1,\cdots,a\}$表示,a为一正整数.

性质(1)称为*马尔可夫性*或*无后效性*.它有极为明显的直观解释.如果把时刻t_k看作现在,那么t_{k+1}是将来的时刻,而$t_1,\cdots,t_{k-1}$则是过去的时刻.马尔可夫性表示在确切知道系统现在的状态的条件下,系统将来的状况与过去的状况无关,或者说为了预测系统将来的状况,只需知道现在系统的状态,更多地了解系统在过去的状况并没有任何帮助.这就是无后效的意思.无论在自然界或社会现象中,都存在着许多随机现象,它们满足或近似地满足这一要求,因此就能用马尔可夫过程来描述这些现象,马尔可夫过程也就成为研究这些现象的有力的数学工具.这也正是马尔可夫过程具有广泛应用的原因.

今后我们将频繁地使用条件概率的乘法公式.为方便初学的读者,我们将它列为一条引理.

引理1.1　设A, B, C为三个随机事件,则

$$\mathrm{P}(BC\mid A)=\mathrm{P}(B\mid A)\mathrm{P}(C\mid AB). \tag{2}$$

证　按条件概率的定义,在$\mathrm{P}(AB)>0$时

$$\begin{aligned}\mathrm{P}(BC\mid A)&=\frac{\mathrm{P}(ABC)}{\mathrm{P}(A)}=\frac{\mathrm{P}(AB)}{\mathrm{P}(A)}\frac{\mathrm{P}(ABC)}{\mathrm{P}(AB)}\\&=\mathrm{P}(B\mid A)\mathrm{P}(C\mid AB).\end{aligned}$$

按我们对有关条件概率的等式的约定,(2)式成立.　□

引理 1.2 设 A,B,C 为三个随机事件,则下列两个等式是等价的:

$$\mathrm{P}(C|AB)=\mathrm{P}(C|B), \tag{3}$$

$$\mathrm{P}(AC|B)=\mathrm{P}(A|B)\mathrm{P}(C|B). \tag{4}$$

证 (3)⇒(4). 若 $\mathrm{P}(AB)=0$,(4)式两边为 0,故设 $\mathrm{P}(AB)>0$. 由引理 1.1

$$\mathrm{P}(AC|B)=\mathrm{P}(A|B)\mathrm{P}(C|AB)=\mathrm{P}(A|B)\mathrm{P}(C|B).$$

(4)⇒(3). 只需考虑 $\mathrm{P}(AB)>0$ 的情形. 由引理 1.1 及(4)式

$$\mathrm{P}(C|AB)=\frac{\mathrm{P}(AC|B)}{\mathrm{P}(A|B)}=\mathrm{P}(C|B). \qquad \square$$

定理 1.1 设 $\{X_n,n\geqslant 0\}$ 为马尔可夫链,则对任意的 $k,m\geqslant 1$ 及 $t_0<t_1<\cdots<t_k<t_{k+1}<\cdots<t_{k+m}, i_0,i_1,\cdots,i_{k+m}\in E$,

$$\begin{aligned}&\mathrm{P}(X_{t_{k+1}}=i_{k+1},\cdots,X_{t_{k+m}}=i_{k+m}\mid X_{t_0}=i_0,\cdots,X_{t_k}=i_k)\\&=\mathrm{P}(X_{t_{k+1}}=i_{k+1},\cdots,X_{t_{k+m}}=i_{k+m}\mid X_{t_k}=i_k).\end{aligned} \tag{5}$$

证 对 m 进行归纳证明. $m=1$ 时(5)式由(1)式即得. 现在设(5)式对 $m-1$ 成立. 记 $A=\{X_{t_0}=i_0,\cdots,X_{t_{k-1}}=i_{k-1}\}$, $B=\{X_{t_k}=i_k\}$, $C_1=\{X_{t_{k+1}}=i_{k+1}\}$, $C_2=\{X_{t_{k+m}}=i_{k+m},\cdots,X_{t_{k+2}}=i_{k+2}\}$,则(5)式的左边及右边分别为 $\mathrm{P}(C_1C_2|AB)$ 及 $\mathrm{P}(C_1C_2|B)$. 利用引理 1.1,引理 1.2 及归纳法假设,我们有

$$\mathrm{P}(C_1C_2|AB)=\mathrm{P}(C_1|AB)\mathrm{P}(C_2|ABC_1)=\mathrm{P}(C_1|B)\mathrm{P}(C_2|C_1). \tag{6}$$

类似地我们也有

$$\mathrm{P}(C_1C_2|B)=\mathrm{P}(C_1|B)\mathrm{P}(C_2|BC_1)=\mathrm{P}(C_1|B)\mathrm{P}(C_2|C_1). \tag{7}$$

比较(6)及(7)即知(5)式对 m 也成立. $\square$

推论 设 $\{X_n,n\geqslant 0\}$ 为一列只取非负整数值的随机变量. $\{X_n,n\geqslant 0\}$ 为马尔可夫链的充要条件为对任意的 $k,m\geqslant 1$,及 $t_0<t_1<\cdots<t_k<t_{k+1}<\cdots<t_{k+m}, i_0,\cdots,i_k,i_{k+1},\cdots,i_{k+m}\in E$,

$$\begin{aligned}&\mathrm{P}(X_{t_0}=i_0,\cdots,X_{t_{k-1}}=i_{k-1},X_{t_{k+1}}=i_{k+1},\cdots,X_{t_{k+m}}=i_{k+m}\mid X_{t_k}=i_k)\\&=\mathrm{P}(X_{t_0}=i_0,\cdots,X_{t_{k-1}}=i_{k-1}\mid X_{t_k}=i_k)\\&\quad\cdot\mathrm{P}(X_{t_{k+1}}=i_{k+1},\cdots,X_{t_{k+m}}=i_{k+m}\mid X_{t_k}=i_k).\end{aligned} \tag{8}$$

证 由定理 1.1 及引理 1.2 即得. $\square$

这推论表明,马尔可夫性可等价地表述为:在确切知道系统现在的状态的条件下,系统的过去与将来相互(条件)独立. 特别可知,在马尔可夫性中,将来与过去处于对称的地位. 因此,若 $\{X_0,X_1,\cdots,X_a\}$ 构成马尔可夫链,则把时间倒过来所得的 $\{X_a,X_{a-1},\cdots,X_0\}$ 也构成马尔可夫链.

(5)式可以写成向量的形式:

$$\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})=(i_{k+1},\cdots,i_{k+m})|$$

$$(X_{t_0},\cdots,X_{t_{k-1}})=(i_0,\cdots,i_{k-1}),X_{t_k}=i_k)$$

$$=\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})=(i_{k+1},\cdots,i_{k+m})\mid X_{t_k}=i_k),\tag{9}$$

这里向量$(i_0,\cdots,i_{k-1})$及$(i_{k+1},\cdots,i_{k+m})$分别是 k 次乘积空间$E^k=\underbrace{E\times\cdots\times E}_{k个}$及 m 次乘积空间$E^m=\underbrace{E\times\cdots\times E}_{m个}$中的一个点.这样就很自然地可以想到把(9)式再作下述进一步的推广.

定理 1.2 设$\{X_n,n\geqslant 0\}$为马尔可夫链,则对任意的 $k,m\geqslant 1,t_0<t_1<\cdots<t_k<t_{k+1}<\cdots<t_{k+m},i\in E$ 及E^k 的子集A,E^m 的子集B,

$$\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})\in B\mid (X_{t_0},\cdots,X_{t_{k-1}})\in A,X_{t_k}=i)$$

$$=\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})\in B\mid X_{t_k}=i).\tag{10}$$

证 由引理 1.2,(10)式等价于下式:

$$\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})\in B,(X_{t_0},\cdots,X_{t_{k-1}})\in A\mid X_{t_k}=i)$$

$$=\mathrm{P}((X_{t_0},\cdots,X_{t_{k-1}})\in A\mid X_{t_k}=i)\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})\in B\mid X_{t_k}=i).\tag{11}$$

我们只要证(11)式,为此把(11)式的左边按状态向量展开,并利用(8)式即可:

$$\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})\in B,(X_{t_0},\cdots,X_{t_{k-1}})\in A\mid X_{t_k}=i)$$

$$=\sum_{(i_1,\cdots,i_m)\in B}\sum_{(j_0,\cdots,j_{k-1})\in A}\mathrm{P}((X_{t_0},\cdots,X_{t_{k-1}})=(j_0,\cdots,j_{k-1}),$$

$$(X_{t_{k+1}},\cdots,X_{t_{k+m}})=(i_1,\cdots,i_m)\mid X_{t_k}=i)$$

$$=\sum_{(i_1,\cdots,i_m)\in B}\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})=(i_1,\cdots,i_m)\mid X_{t_k}=i)$$

$$\cdot\sum_{(j_0,\cdots,j_{k-1})\in A}\mathrm{P}((X_{t_0},\cdots,X_{t_{k-1}})=(j_0,\cdots,j_{k-1})\mid X_{t_k}=i)$$

$$=\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})\in B\mid X_{t_k}=i)\mathrm{P}((X_{t_0},\cdots,X_{t_{k-1}})\in A\mid X_{t_k}=i).\quad\square$$

我们把马尔可夫性从(1)推广到了(10).马尔可夫性的这一推广使它的直观意义更明朗了.在(10)式中把时刻 t_k 作为现在,那么$\{(X_{t_0},\cdots,X_{t_{k-1}})\in A\}$就是过去的事件,$\{(X_{t_{k+1}},\cdots,X_{t_{k+m}})\in B\}$就是将来的事件.在已知现在的状态的条件下,将来的事件与过去的事件独立.这里将来的事件$\{(X_{t_{k+1}},\cdots,X_{t_{k+m}})\in B\}$不只涉及一个将来的时刻,可以涉及任意有限个将来的时刻.再进一步,由概率的连续性(即可列可加性)可知,将来的事件也可以涉及系统在无穷多个将来时刻的状况,但(10)式依然成立.由于涉及无穷多个时刻的事件难以给出一个统一的表达式,所以这里我们不写出具体的式子.在今后的讨论中我们将会看到各种形式的涉及无穷多个时刻的事件,那时我们仍运用马尔可夫性而不再特别指明.我们还要特别指出,如果在(10)式中还想把$\{X_{i_k}=i\}$换成$\{X_{i_k}\in D\}$,D 为状态空间E 的一个子集,它包含的状态多于一个,那么一般说来,

$$\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})\in B\mid (X_0,\cdots,X_{t_{k-1}})\in A,X_{t_k}\in D)$$

$$=\mathrm{P}((X_{t_{k+1}},\cdots,X_{t_{k+m}})\in B\mid X_{t_k}\in D)$$

就不再成立了(参阅下面的例 1.1).换句话说,在马尔可夫性中要求确切地知道现在的状态,不能只含糊地知道现在的状态是某些状态中的一个.这个要求也是十分自然的.

前面的讨论使我们扩充了应用马尔可夫性的范围.但为了验证马尔可夫性,我们希望尽可能只要验证形式上更简单些的要求.对于离散时间的情形,我们有下述定理满足这个要求.

定理 1.3 设$\{X_n, n\geqslant 0\}$为一列只取非负整数值的随机变量.若对任意的 $n\geqslant 1$ 及非负整数 $i_0, i_1, \cdots, i_n, i_{n+1}$,

$$\mathrm{P}(X_{n+1}=i_{n+1} \mid X_0=i_0, X_1=i_1, \cdots, X_n=i_n)=\mathrm{P}(X_{n+1}=i_{n+1} \mid X_n=i_n), \tag{12}$$

则$\{X_n, n\geqslant 0\}$为马尔可夫链.

证 我们首先证明:对任意的 $n\geqslant 1$ 及 $k<n$,

$$\mathrm{P}(X_{n+1}=i_{n+1} \mid X_k=i_k, \cdots, X_n=i_n)=\mathrm{P}(X_{n+1}=i_{n+1} \mid X_n=i_n). \tag{13}$$

或由引理 1.2,证明(13)的等价式:

$$\begin{aligned}&\mathrm{P}(X_k=i_k, \cdots, X_{n-1}=i_{n-1}, X_{n+1}=i_{n+1} \mid X_n=i_n)\\&\quad=\mathrm{P}(X_k=i_k, \cdots, X_{n-1}=i_{n-1} \mid X_n=i_n)\mathrm{P}(X_{n+1}=i_{n+1} \mid X_n=i_n).\end{aligned} \tag{14}$$

而(12)式等价于

$$\begin{aligned}&\mathrm{P}(X_0=i_0, \cdots, X_{n-1}=i_{n-1}, X_{n+1}=i_{n+1} \mid X_n=i_n)\\&\quad=\mathrm{P}(X_0=i_0, \cdots, X_{n-1}=i_{n-1} \mid X_n=i_n)\mathrm{P}(X_{n+1}=i_{n+1} \mid X_n=i_n),\end{aligned}$$

由此式推导(14)式的过程如下:

$$\begin{aligned}&\mathrm{P}(X_k=i_k, \cdots, X_{n-1}=i_{n-1}, X_{n+1}=i_{n+1} \mid X_n=i_n)\\&\quad=\sum_{j_0, \cdots, j_{k-1}} \mathrm{P}(X_0=j_0, \cdots, X_{k-1}=j_{k-1}, X_k=i_k, \cdots, X_{n-1}=i_{n-1},\\&\qquad\qquad X_{n+1}=i_{n+1} \mid X_n=i_n)\\&\quad=\sum_{j_0, \cdots, j_{k-1}} \mathrm{P}(X_0=j_0, \cdots, X_{k-1}=j_{k-1}, X_k=i_k, \cdots, X_{n-1}=i_{n-1} \mid X_n=i_n)\\&\qquad\cdot\mathrm{P}(X_{n+1}=i_{n+1} \mid X_n=j_n)\\&\quad=\mathrm{P}(X_k=i_k, \cdots, X_{n-1}=i_{n-1} \mid X_n=i_n)\mathrm{P}(X_{n+1}=i_{n+1} \mid X_n=i_n).\end{aligned}$$

有了(13)式,照搬定理 1.1 的证明(取 $t_l=l, l=0,1,\cdots,k+m$)可得:对任意的 $k, m\geqslant 1$ 及 $i_0, i_1, \cdots, i_{k+m}\in E$,

$$\begin{aligned}&\mathrm{P}(X_{k+1}=i_{k+1}, \cdots, X_{k+m}=i_{k+m} \mid X_0=i_0, \cdots, X_k=i_k)\\&\quad=\mathrm{P}(X_{k+1}=i_{k+1}, \cdots, X_{k+m}=i_{k+m} \mid X_k=i_k).\end{aligned}$$

再照搬定理 1.2 的证明(以 n 代替 k)可得:对任意的 $n, m\geqslant 1, i\in E$ 及 E^n 的子集 A, E^m 的子集 B,

$$\begin{aligned}&\mathrm{P}((X_{n+1}, \cdots, X_{n+m})\in B, (X_0, \cdots, X_{n-1})\in A, X_n=i)\\&\quad=\mathrm{P}((X_{n+1}, \cdots, X_{n+m})\in B \mid X_n=i).\end{aligned} \tag{15}$$

为了从(15)式得到(1)式,假设 $t_k=n,t_{k+1}=n+m$,只要在(15)中取

$$A=\{(j_0,\cdots,j_{n-1})\in E^n:j_{t_l}=i_l,l=0,\cdots,k-1\},$$
$$B=\{(j_1,\cdots,j_m)\in E^m:\ j_m=i_{k+1}\}$$

就可以了. □

从前面的讨论中已经可以看出,条件概率 $\mathrm{P}(X_{n+m}=j\mid X_m=i)$起到重要的作用.为了得到更鲜明的直观形象,我们称状态的变化为转移,每一个单位时间移动一步.所以条件概率 $\mathrm{P}(X_{n+m}=j\mid X_m=i)$称为在时刻 m 从状态 i 出发,经过 n 步转移到状态 j 的转移概率.如果对一切 $n\geqslant1,m\geqslant0$ 及 $i,j\in E$,

$$\mathrm{P}(X_{n+m}=j\mid X_m=i)$$

与 m 无关,记为 $p_{ij}^{(n)}$,则称这马尔可夫链为(时间)齐次的或具有平稳的转移概率.今后我们只讨论齐次马尔可夫链,因此将省略齐次两字.$p_{ij}^{(n)}$称为 n 步(或 n 阶)转移概率.特别记

$$p_{ij}^{(0)}=\delta_{ij}=\begin{cases}1, & i=j,\\0, & i\neq j,\end{cases}\quad p_{ij}^{(1)}=p_{ij}.$$

p_{ij}就称为由状态 i 到状态 j 的转移概率.

定理 1.4 对任意的 $n,m\geqslant0$ 及 $i,j\in E$,

$$p_{ij}^{(n+m)}=\sum_k p_{ik}^{(n)}p_{kj}^{(m)}. \tag{16}$$

证 按时刻 n 的状态进行分解,再用马尔可夫性有

$$\begin{aligned}p_{ij}^{(n+m)}&=\mathrm{P}(X_{n+m}=j\mid X_0=i)\\&=\sum_k\mathrm{P}(X_{n+m}=j,X_n=k\mid X_0=i)\\&=\sum_k\mathrm{P}(X_n=k\mid X_0=i)\mathrm{P}(X_{n+m}=j\mid X_n=k,X_0=i)\\&=\sum_k\mathrm{P}(X_n=k\mid X_0=i)\mathrm{P}(X_{n+m}=j\mid X_n=k)=\sum_k p_{ik}^{(n)}p_{kj}^{(m)}.\end{aligned}$$ □

(16)式称为科尔莫戈罗夫(Kolmogorov)-查普曼(Chapman)方程,它的直观意义也是十分明显的.从状态 i 出发经过 $n+m$ 步到达状态 j 可分成两阶段走:先从状态 i 出发经过 n 步到达状态 k,然后再从状态 k 出发经过 m 步到达状态 j.由马尔可夫性,后一阶段的状态转移与前一阶段的状态转移独立,故而两个阶段的转移概率应相乘.经过 n 步所能达到的状态 k 不受任何限制,因此应对全部的 k 求和.(16)式的证明虽然不长,但它的证明步骤对于研究马尔可夫链来讲,却是十分典型的.首先,按照某种方式将事件作分解(这里是按时刻 n 的状态作分解),然后相继地利用条件概率的乘法公式、马尔可夫性及齐次性.今后我们可能采用不同的分解事件的方法,然而都离不开这一基本模式.读者应充分注意,并学会运用这一证明方法.充分地明了公式的直观意义,也有助于掌握这个证明方法,读者也应尽可能地理解所遇到的每一个公式的直观意义.

如果把转移概率写成矩阵的形式:

$$\mathbf{P}^{(n)}=(p_{ij}^{(n)})_{i,j\in E},$$

那么科尔莫戈罗夫－查普曼方程具有下列简单的形式：

$$\mathbf{P}^{(n+m)}=\mathbf{P}^{(n)}\mathbf{P}^{(m)},\qquad n,m\geqslant 0. \tag{17}$$

只是这里的矩阵可以是无穷的，但乘法规则却与有限矩阵一样，可按行乘列法则定义．$\mathbf{P}=(p_{ij})$ 称为马尔可夫链的转移概率矩阵，它满足下列条件：

(1) $p_{ij}\geqslant 0,\ i,j\in E$；

(2) $\sum\limits_{j} p_{ij}=1,\ i\in E$.

（满足这两个条件的矩阵称为随机矩阵．）$\mathbf{P}^{(n)}=\mathbf{P}^{n}$，$n\geqslant 1$ 称为 n 步转移概率矩阵．易见，$\mathbf{P}^{(0)}=(\delta_{ij})$是单位阵．

初始状态 X_0 的概率分布

$$\mathsf{P}(X_0=i)=p_i,\qquad i\in E$$

称为马尔可夫链的*初始分布*．不难用初始分布及转移概率写出马尔可夫链的有限维分布．首先利用前面已强调指明的证明方法可以得到

$$\begin{aligned}
&\mathsf{P}(X_1=i_1,\cdots,X_n=i_n\mid X_0=i_0)\\
&\quad=\mathsf{P}(X_1=i_1\mid X_0=i_0)\mathsf{P}(X_2=i_2\mid X_0=i_0,X_1=i_1)\\
&\qquad\cdots\mathsf{P}(X_n=i_n\mid X_0=i_0,\cdots,X_{n-1}=i_{n-1})\\
&\quad=\mathsf{P}(X_1=i_1\mid X_0=i_0)\mathsf{P}(X_2=i_2\mid X_1=i_1)\\
&\qquad\cdots\mathsf{P}(X_n=i_n\mid X_{n-1}=i_{n-1})\\
&\quad=p_{i_0i_1}p_{i_1i_2}\cdots p_{i_{n-1}i_n}.
\end{aligned}\tag{18}$$

于是有限维分布为：

$$\mathsf{P}(X_0=i_0,X_1=i_1,\cdots,X_n=i_n)=p_{i_0}p_{i_0i_1}p_{i_1i_2}\cdots p_{i_{n-1}i_n}. \tag{19}$$

反过来，如果一个随机变量序列$\{X_n,n\geqslant 0\}$的有限维分布由(19)式给出，其中$\{p_i,i\in E\}$为一概率分布，(p_{ij})为一随机矩阵，那么由定理 1.3 就可知道$\{X_n,n\geqslant 0\}$为马尔可夫链，即(12)式成立，且(p_{ij})就是它的转移概率矩阵，$\{p_i\}$是它的初始分布．因此，马尔可夫链就是有限维分布的形式为(19)的随机变量序列．

对于马尔可夫链，我们主要是研究它的状态如何随时间的推移而转移，即研究状态转移的概率规律．说得更明确些，就是要研究完全由转移概率矩阵所决定的性质．对初始分布一般不作特别的规定，允许它取任意的概率分布．因此，从数学上看，就可以把一个随机矩阵当作一个马尔可夫链，这个随机矩阵就是这个马尔可夫链的转移概率矩阵，而 n 步转移概率矩阵则由(19)式确定．

与(18)式的证明一样，我们有

$$\mathsf{P}(X_{n+1}=i_1,\cdots,X_{n+m}=i_m\mid X_n=i_0)=p_{i_0i_1}p_{i_1i_2}\cdots p_{i_{m-1}i_m}$$

与 n 无关．若 B 为$\underbrace{E\times\cdots\times E}_{m个}$的一个子集，则

$$\mathsf{P}((X_{n+1},\cdots,X_{n+m})\in B\mid X_n=i_0)=\mathsf{P}((X_1,\cdots,X_m)\in B\mid X_0=i_0) \tag{20}$$

同样与 n 无关.由概率的连续性(可列可加性)也可把事件 $\{(X_{n+1},\cdots,X_{n+m})\in B\}$ 换成涉及系统在时刻 n 之后的无穷多个时刻状况的事件,相应的(20)式依然成立.因此,在已知系统在时刻 n 所处的状态的条件下,讨论系统在时刻 n 之后的状况时,可以把时刻 n 看作初始时刻,其它的时间参数作相应的推移.这就是齐次性假定的意义.今后我们也要经常这样地运用齐次性,不再特别指明了.

例 1.1 无限制的随机游动 随机游动是指一个质点在直线上的某个范围内随机地逐步移动,也可以讨论在平面上或空间中的随机游动.这里我们先讨论直线上的随机游动.假定质点的位置总在整数点上,每隔一个单位时间移动一步,可以移动到左、右相邻的位置,或停留在原来的位置上,且转移的概率只与该时刻质点的位置有关,与该时刻以及质点在以前的位置无关.以 X_n 表示质点在时刻 n 的位置,则 $\{X_n,n\geqslant 0\}$ 就是一个马尔可夫链,称为*直线上的随机游动*.

如果随机游动的状态空间为 $\mathbb{Z}=\{\cdots,-1,0,1,\cdots\}$,转移概率为

$$p_{i,i+1}=p,\quad p_{i,i-1}=q=1-p,\quad 0<p<1,\quad i\in\mathbb{Z},$$

那么称这样的随机游动为 *(p,q) 随机游动*.这时游动没有任何范围的限制.如果 $p=q=1/2$.则称这随机游动是*对称的*.容易看出,每游动一步可看为作一次有两个结果的试验,这些试验是相互独立的.因此利用独立试验序列的熟知的结果,我们不难计算 n 步转移概率 $p_{ij}^{(n)}$.从 i 经过 n 步到 j,必须向右移动的步数比向左移动的步数多 $j-i$ 步.因此这 n 步中必须向右移动 $(n+j-i)/2$ 步,向左移动 $[n-(j-i)]/2$ 步,$n+j-i$ 必须是偶数.由此可得

$$p_{ij}^{(n)}=\begin{cases}C_n^{(n+j-i)/2}p^{(n+j-i)/2}q^{(n-j+i)/2}, & n+j-i\ \text{为偶数},\\ 0, & n+j-i\ \text{为奇数}.\end{cases}$$

随机游动除了本身具有一定的物理意义外,它实际上是一个可以描述多种随机现象的数学模型.许多随机现象都能归结成各种形式的随机游动,从而引起了研究随机游动的兴趣,这情况就像在概率论中人们热衷于讨论从罐子中摸球的模型一样.此外,随机游动又是形式最为简单的马尔可夫链,把它作为例子作详尽的讨论也是自然的.

现在取初始分布为 $\mathsf{P}(X_0=0)=\mathsf{P}(X_0=2)=1/2$, $D=\{1,3\}$,这时

$$\begin{aligned}\mathsf{P}(X_1=1)&=\mathsf{P}(X_0=0)\mathsf{P}(X_1=1\mid X_0=0)+\mathsf{P}(X_0=2)\mathsf{P}(X_1=1\mid X_0=2)\\&=\frac{1}{2}p+\frac{1}{2}q=\frac{1}{2},\end{aligned}$$

$$\mathsf{P}(X_1=3)=\mathsf{P}(X_0=2)\mathsf{P}(X_1=3\mid X_0=2)=\frac{1}{2}p,$$

$$\begin{aligned}\mathsf{P}(X_2=2\mid X_1\in D)&=\frac{\mathsf{P}(X_1=1)\mathsf{P}(X_2=2\mid X_1=1)+\mathsf{P}(X_1=3)\mathsf{P}(X_2=2\mid X_1=3)}{\mathsf{P}(X_1=1)+\mathsf{P}(X_1=3)}\\&=\frac{\frac{1}{2}p+\frac{1}{2}pq}{\frac{1}{2}+\frac{1}{2}p}=p\,\frac{1+q}{1+p},\end{aligned}$$

$$\mathsf{P}(X_2=2\mid X_0=0,X_1\in D)=\mathsf{P}(X_2=2\mid X_0=0,X_1=1)=p.$$

因此 $p\neq q$ 时，

$$\mathrm{P}(X_2=2\mid X_0=0,X_1\in D)\neq\mathrm{P}(X_2=2\mid X_1\in D).$$

这表明在马尔可夫性中要求确切地知道现在的状态，不能含糊. □

下面的定理提供了一个非常有用的获得马尔可夫链的模式.

定理 1.5 设整值随机变量序列 $\{X_n,n\geqslant0\}$ 满足下列两个条件：

(1) $n\geqslant1$ 时，$X_n=f(X_{n-1},\xi_n)$；

(2) $\{\xi_n,n\geqslant1\}$ 为独立同分布随机变量，且 X_0 与 $\{\xi_n,n\geqslant1\}$ 也相互独立，则 $\{X_n,n\geqslant0\}$ 是马尔可夫链，其转移概率为

$$p_{ij}=\mathrm{P}(f(i,\xi_1)=j). \tag{21}$$

证 由定理 1.3，我们只要证明(12)式成立. 设 $n\geqslant1$，注意到 ξ_{n+1} 与 $X_0,X_1,\cdots,X_n$ 相互独立，

$$\begin{aligned}\mathrm{P}(X_{n+1}=i_{n+1}\mid X_0=i_0,\cdots,X_n=i_n)&=\mathrm{P}(f(X_n,\xi_{n+1})=i_{n+1}\mid X_0=i_0,\cdots,X_n=i_n)\\&=\mathrm{P}(f(i_n,\xi_{n+1})=i_{n+1}\mid X_0=i_0,\cdots,X_n=i_n)\\&=\mathrm{P}(f(i_n,\xi_{n+1})=i_{n+1}),\end{aligned} \tag{22}$$

同样地，

$$\mathrm{P}(X_{n+1}=i_{n+1}\mid X_n=i_n)=\mathrm{P}(f(i_n,\xi_{n+1})=i_{n+1}). \tag{23}$$

由(22)及(23)即知(12)式成立. 由(23)，转移概率为(21). □

推论 设 $\{\xi_n,n\geqslant0\}$ 为一列独立的整值随机变量，且 $\{\xi_n,n\geqslant1\}$ 同分布：

$$\mathrm{P}(\xi_n=k)=\alpha_k,\quad k\in\mathbb{Z},\quad n\geqslant1,$$

令

$$X_n=\xi_0+\xi_1+\cdots+\xi_n,\qquad n\geqslant0,$$

则 $\{X_n,n\geqslant0\}$ 为马尔可夫链，且转移概率为 $p_{ij}=\alpha_{j-i}$.

证 事实上，在定理 1.5 中取 $f(a,b)=a+b$ 即得. □

这推论中的 $\{X_n,n\geqslant0\}$ 就是我们熟知的具有独立平稳增量的随机序列. 反过来，若马尔可夫链的状态空间为 $\mathbb{Z}$，且转移概率 p_{ij} 只依赖于 $j-i$（这时马尔可夫链称为空间齐次的），那么这个马尔可夫链是具有独立平稳增量的随机序列，事实上记 $\xi_n=X_n-X_{n-1},n\geqslant1,\xi_0=X_0,\alpha_k=p_{i,i+k}$，则 $\sum_k\alpha_k=1$. 由(12)式

$$\begin{aligned}\mathrm{P}(\xi_0=i_0,\xi_1=i_1,\cdots,\xi_n=i_n)&=\mathrm{P}(X_0=i_0,X_1=i_0+i_1,\cdots,X_n=i_0+i_1+\cdots+i_n)\\&=\mathrm{P}(X_0=i_0)p_{i_0,i_0+i_1}\cdots p_{i_0+\cdots+i_{n-1},i_0+\cdots+i_n}\\&=\mathrm{P}(X_0=i_0)\alpha_{i_1}\cdots\alpha_{i_n}.\end{aligned}$$



由此可知

$$\mathrm{P}(\xi_n = k) = a_k, \qquad k \in \mathbb{Z}, \quad n \geqslant 1,$$

$$\mathrm{P}(\xi_0 = i_0, \xi_1 = i_1, \cdots, \xi_n = i_n) = \mathrm{P}(\xi_0 = i_0)\mathrm{P}(\xi_1 = i_1)\cdots\mathrm{P}(\xi_n = i_n).$$

这即表明$\{\xi_n, n \geqslant 0\}$为一列独立随机变量,$\{\xi_n, n \geqslant 1\}$同分布,从而$\{X_n, n \geqslant 0\}$是具有独立平稳增量的随机序列.

容易看出,当

$$\mathrm{P}(\xi_n = 1) = p, \qquad \mathrm{P}(\xi_n = -1) = q, \qquad n \geqslant 1$$

时,我们就得到(p, q)随机游动.

例 1.2 存储问题 设一家电视机商店最多可存放 S 台电视机.开始时商店进货进足 S 台电视机.若在第 n 个月中顾客欲购的电视机台数(需求量)为 ξ_n,第 n 个月底盘点时所剩的电视机台数记为 X_n.盘点后决定是否进货.决策的方法如下:若 $X_n \leqslant s$,就立即进货至 S 台,若 $X_n > s$,则不进货.假设$\{\xi_n, n \geqslant 1\}$为 i.i.d. 随机变量序列,其共同分布为$\{q_k, k \geqslant 0\}$.这时$\{X_n, n \geqslant 0\}$是马尔可夫链.事实上,我们有

$$X_0 = S,$$

$$X_n = \begin{cases} \max(0, X_{n-1} - \xi_n), & \text{当 } s < X_{n-1} \leqslant S, \\ \max(0, S - \xi_n), & \text{当 } X_{n-1} \leqslant s, \end{cases} \qquad n \geqslant 1.$$

如果我们定义一个函数:

$$f(a, b) = \begin{cases} \max(0, a - b), & \text{当 } s < a \leqslant S, \\ \max(0, S - b), & \text{当 } a \leqslant s, \end{cases}$$

那么$\{X_n, n \geqslant 0\}$满足定理 1.5 的条件,它是状态空间为$\{0, 1, \cdots, S\}$的马尔可夫链,转移概率为

$$p_{ij} = \begin{cases} a_s, & j = 0,\ i \leqslant s, \\ a_i, & j = 0,\ i > s, \\ q_{s-j}, & 0 < j \leqslant S,\ i \leqslant s, \\ q_{i-j}, & 0 < j \leqslant i,\ i > s, \\ 0, & \text{其它}, \end{cases}$$

其中 $a_i = \sum_{j=i}^{\infty} q_j$.

我们所讨论的是一个采取(s, S)策略的存储模型.商店感兴趣的是如何决定 s 的大小.从加快资金的流动、减少存储保管的费用等因素出发,s 应取得大些,但 s 大,商品就会脱销,减少营业额与利润,对商店也不利.所以要根据实际所测得的各种因素的数据及利弊的大小,综合地决定 s 的值.显然,掌握马尔可夫链$\{X_n, n \geqslant 0\}$的性质总是十分必要的. □

例 1.3 我们考察在某个容器内某种物质的分子个数的变迁.设在单位时间内,容器中每一个分子以概率 p 留在容器内,以概率 $q = 1 - p$ 逸出.同时又有新的分子进入容器,进入的分子数服从参数为 λ 的泊松分布.我们假设各个分子的运动是相互独立地进行的.以 X_n 记时刻 n 容器中的分子数.由上面的假定容易看出,$\{X_n, n \geqslant 0\}$是马尔可夫链.因为下一个时刻的分子数只取

决于现在容器中的分子及新进入的分子数.这些都与以前容器中分子的个数无关.

我们来计算转移概率 p_{ij}.设容器中原有的分子中有 k 个留下,则有 $i-k$ 个逸出.新进入的分子数为 $j-k$.容易看到,应有 $k\leqslant i$ 及 $k\leqslant j$,所以 k 的变化范围是从 0 到 $\min(i,j)$.由各分子运动为相互独立的假定,我们得到:

$$p_{ij}=\mathrm{e}^{-\lambda}\sum_{k=0}^{\min(i,j)}\mathrm{C}_i^k p^k q^{i-k}\frac{\lambda^{j-k}}{(j-k)!}. \qquad \square$$

例 1.4 设 $\{Y_n,n\geqslant 0\}$ 为马尔可夫链 $\{X_n,n\geqslant 0\}$ 的函数:

$$Y_n=f(X_n),\qquad n\geqslant 0,$$

其中 f 为从 E_X 到 E_Y 的函数,E_X 为 $\{X_n,n\geqslant 0\}$ 的状态空间,E_Y 为 $\{Y_n,n\geqslant 0\}$ 的状态空间.一般说来,$\{Y_n,n\geqslant 0\}$ 不再是马尔可夫链.若对任一 $k\in E_Y$,存在定义在 E_Y 上的函数 h_k,使得对一切 $i\in E_X$,有

$$\sum_{\{j:f(j)=k\}}p_{ij}=h_k(f(i)),\tag{24}$$

则 $\{Y_n,n\geqslant 0\}$ 是马尔可夫链.事实上,对一切 $n\geqslant 1,j_0,j_1,\cdots,j_n\in E_Y$,

$$\begin{aligned}
&\mathrm{P}(Y_0=j_0,Y_1=j_1,\cdots,Y_n=j_n)\\
&=\sum_{\{(i_0,\cdots,i_n):f(i_l)=j_l,0\leqslant l\leqslant n\}}p_{i_0}p_{i_0i_1}\cdots p_{i_{n-1}i_n}\\
&=\sum_{\{(i_0,\cdots,i_{n-1}):f(i_l)=j_l,0\leqslant l\leqslant n-1\}}p_{i_0}p_{i_0i_1}\cdots p_{i_{n-2}i_{n-1}}\sum_{\{i_n:f(i_n)=j_n\}}p_{i_{n-1}i_n}\\
&=h_{j_n}(j_{n-1})\sum_{\{(i_0,\cdots,i_{n-1}):f(i_l)=j_l,0\leqslant l\leqslant n-1\}}p_{i_0}p_{i_0i_1}\cdots p_{i_{n-2}i_{n-1}}\\
&=h_{j_n}(j_{n-1})\cdots h_{j_2}(j_1)h_{j_1}(j_0)\mathrm{P}(Y_0=j_0).
\end{aligned}$$

因此 $\{Y_n,n\geqslant 0\}$ 为马尔可夫链,且其转移概率为

$$\mathrm{P}(Y_{n+1}=k\mid Y_n=j)=h_k(j).$$

反之,如果不论 X_0 的分布如何,$\{Y_n,n\geqslant 0\}$ 总为具有同一个转移概率矩阵的马尔可夫链,那么条件(24)必定成立.这一事实留给有兴趣的读者自行证明.

现在设马尔可夫链 $\{X_n,n\geqslant 0\}$ 的状态空间为 $\mathbb{Z}$,其转移概率满足条件:

$$p_{-i,-j}=p_{ij},\qquad i,j\in\mathbb{Z}.\tag{25}$$

这时 $\{|X_n|,n\geqslant 0\}$ 为马尔可夫链.事实上,这时 $E_X=\mathbb{Z}$,$Y_n=f(X_n),n\geqslant 0,f(i)=|i|,i\in\mathbb{Z}$,$E_Y=\mathbb{N}$.令

$$h_k(j)=\begin{cases}p_{jk}+p_{j,-k}, & k\geqslant 1,\\ p_{j0}, & k=0,\end{cases}\qquad j\in\mathbb{N}.$$

由条件(25)可知条件(24)成立. $\square$

习　题

1-1. 设从数字 $1,2,\cdots,a$ 中任取一数为 X_0,$n\geqslant 1$ 时,则从 $1,2,\cdots,X_{n-1}$ 中任取一数为 X_n,

那么$\{X_n, n\geqslant 0\}$为马尔可夫链,试写出它的转移概率矩阵.

1-2. 设连续地掷一个均匀的硬币.定义$X_n=0$,若第n次掷得正面;$X_n=k$,若第$n, n-1, \cdots, n-k+1$次掷得反面,而第$n-k$次掷得正面.这里我们规定第0次掷得正面,那么$\{X_n, n\geqslant 0\}$为马尔可夫链,试写出它的转移概率矩阵.

1-3. 一个"传染病"模型如下:设a个人中某些人已患流行性感冒.假定:(1) 当一个病人遇见一个健康者时,后者被传染的概率为α;(2) 所有的接触都是两人之间的接触;(3) 一切成对的接触都是等可能的;(4) 在每个单位时间内只发生一次接触.以X_n记时刻n患病的人数,则$\{X_1, \cdots, X_a\}$为马尔可夫链,试写出它的转移概率矩阵.

1-4. 考察汽车单向通过某路口的情形.假设一秒钟可通过一辆汽车.路口设置有自动红绿灯.红灯持续r秒,绿灯持续g秒.以"红灯—绿灯"为一个单元时间.在第n个单元开始时正在等候的汽车数记为X_{n-1},在这单元中来到的汽车数记为ξ_n.设$\{\xi_n, n\geqslant 1\}$为i.i.d.随机变量,其共同分布为$\{q_k, k=0,1,\cdots,r+g\}$,而开始时路口没有汽车在等候.那么$\{X_n, n\geqslant 0\}$为马尔可夫链,试写出它的转移概率矩阵.

1-5. 设某一水库的容量为c立方米.以X_n记时刻n水库的储水量.在时间区间$[n, n+1)$内流入水库的水量为ξ_{n+1},超过水库容量的水即溢出.在时间区间$[n, n+1)$末从水库放掉m $(<c)$立方米水(其中包括库满而溢出的部分,如果溢出的水量已超过m立方米就不再放水).假设$\{\xi_n, n\geqslant 1\}$i.i.d.,且与X_0独立,$\mathrm{P}(\xi_n=k)=\alpha_k, k\geqslant 0$.那么$\{X_n, n\geqslant 0\}$为马尔可夫链,试写出它的转移概率矩阵.

1-6. 设马尔可夫链的转移概率为

$$(1)\ \mathbf{P}=\begin{pmatrix}0 & 1/2 & 1/2\\ 1 & 0 & 0\\ 1 & 0 & 0\end{pmatrix};\qquad (2)\ \mathbf{P}=\begin{pmatrix}1 & 0 & 0\\ 1/4 & 1/2 & 1/4\\ 0 & 0 & 1\end{pmatrix}.$$

求$\mathbf{P}^{(n)}$.

1-7. 设只有两个状态$\{0,1\}$的马尔可夫链的转移概率矩阵为

$$\mathbf{P}=\begin{pmatrix}q_0 & 1-q_0\\ 1-q_1 & q_1\end{pmatrix},\quad 0<q_0+q_1<2.$$

证明:n步转移概率矩阵为

$$\mathbf{P}^{(n)}=\frac{1}{2-q_0-q_1}\begin{pmatrix}1-q_1 & 1-q_0\\ 1-q_1 & 1-q_0\end{pmatrix}+\frac{(q_0+q_1-1)^n}{2-q_0-q_1}\begin{pmatrix}1-q_0 & -(1-q_0)\\ -(1-q_1) & 1-q_1\end{pmatrix}.$$

由此求$\lim_{n\to\infty}\mathbf{P}^{(n)}$.

1-8. 设马尔可夫链的转移概率为

$$p_{ij}=\begin{cases}1/[(i+1)(i+2)], & j<i+1,\\ (i+1)/(i+2), & j=i+1,\\ 0, & j>i+1.\end{cases}$$

证明：n 步转移概率为

$$p_{ij}^{(n)}=\begin{cases}n/[(i+n)(i+n+1)], & j<i+n,\\ (i+1)/(i+n+1), & j=i+n,\\ 0, & j>i+n.\end{cases}$$

1-9. 设 $\{\xi_n, n\geqslant 1\}$ i.i.d.，且

$$\mathrm{P}(\xi_n=1)=\mathrm{P}(\xi_n=-1)=1/2.$$

令

$$S_0=0,\ S_n=\sum_{k=1}^{n}\xi_k, n\geqslant 1;\quad M_n=\max_{0\leqslant k\leqslant n}S_k, n\geqslant 0.$$

证明：$\{|S_n|, n\geqslant 0\}$ 及 $\{M_n-S_n, n\geqslant 0\}$ 为马尔可夫链.

1-10. 设马尔可夫链 $\{X_n, n\geqslant 0\}$ 的转移概率矩阵为 (p_{ij})，初始分布为 $\{p_k, k\geqslant 0\}$. 又设 $\{U_n, n\geqslant 0\}$ 为 i.i.d. 随机变量序列，其共同分布为 $(0,1]$ 上的均匀分布. 令

$$f(i,u)=\sum_{k=0}^{\infty}k1\Big(\sum_{j=0}^{k-1}p_{ij},\sum_{j=0}^{k}p_{ij}\Big](u),\qquad i\in E, u\in(0,1],$$

$$Y_0=\sum_{k=0}^{\infty}k1\Big(\sum_{j=0}^{k-1}p_j,\sum_{j=0}^{k}p_{ij}\Big](U_0),\qquad Y_{n+1}=f(Y_n,U_{n+1}), n\geqslant 0,$$

则 $\{Y_n, n\geqslant 0\}$ 为马尔可夫链，且与 $\{X_n, n\geqslant 0\}$ 同分布. 本题可看作为定理 1.5 的逆定理.

§2.2 状态的分类与周期

今后我们将总是对一个固定的齐次马尔可夫链进行讨论，不再重复指明.

定义 如果存在 $n\geqslant 1$，使 $p_{ij}^{(n)}>0$，则称从状态 i 可到达状态 j，记作 $i\to j$. 反之，以 $i\not\to j$ 表示从状态 i 不能到达状态 j，即对一切 $n\geqslant 1$，$p_{ij}^{(n)}=0$.

如果 $i\to j$，且 $j\to i$，则称状态 i 与状态 j 相通，记作 $i\leftrightarrow j$.

定理 2.1 若 $i\to j$，$j\to k$，则 $i\to k$.

证 因为 $i\to j$，$j\to k$，必有 $n\geqslant 1$，$m\geqslant 1$ 使得

$$p_{ij}^{(n)}>0,\qquad p_{jk}^{(m)}>0.$$

由科尔莫戈罗夫-查普曼方程

$$p_{ik}^{(n+m)}=\sum_{l}p_{il}^{(n)}p_{lk}^{(m)}\geqslant p_{ij}^{(n)}p_{jk}^{(m)}>0,$$

而 $n+m\geqslant 1$，所以 $i\to k$. □

两个状态相通的关系显然是对称的. 由定理 2.1，这关系也是传递的：由 $i\leftrightarrow j$ 及 $j\leftrightarrow k$ 可得 $i\leftrightarrow k$. 因此可按此关系把状态进行分类，使相通的状态归属于同一个类. 整个状态空间就可分解成若干个类. 每一个状态属于且只属于其中的一个类. 包含状态 i 的类记作 $C(i)$，则

$$C(i)=\{i\}\cup\{j\neq i: j\leftrightarrow i\}.$$

如果状态 i 不与任何别的状态相通,那么它自身就构成一个类:$C(i)=\{i\}$.

由定理2.1,若 $i\to j$,则由 $C(i)$中的任何一个状态也可到达 $C(j)$中的任何一个状态,这时也记为 $C(i)\to C(j)$.

定义 状态 i 称为本质的,若 $i\to j$ 时,必有 $j\to i$.反之,若有状态 j 使得 $i\to j$,但 $j\not\to i$,则称 i 为非本质的.

定理2.2 若状态 i 是本质的,$i\to j$,则状态 j 也是本质的.

证 若 $j\to k$,因为 $i\to j$,所以 $i\to k$,而 i 是本质状态,故 $k\to i$.再由 $i\to j$ 得到 $k\to j$.这即表明 j 是本质状态. □

由定理2.2,同一个类中的状态或全为本质的,或全为非本质的,相应地我们就称这个类为本质类或非本质类.再由类的定义可知,从本质类中的状态不能到达别的类中的状态,但从非本质类的状态就有可能到达本质状态类.本质类的这个性质导致下述闭集的概念.

定义 一个状态的集合 A 称为闭集,如果对每个 $i\in A$

$$\sum_{j\in A} p_{ij}=1. \tag{1}$$

闭集 A 称为极小的,若 A 的任何一个真子集不是闭集.

用归纳法不难证明,若 A 为闭集,对任意 $n\geqslant 1$ 及 $i\in A$ 有

$$\sum_{j\in A} p_{ij}^{(n)}=1. \tag{2}$$

事实上,$n=1$ 时(2)式即(1)式.若(2)式对 $n-1$ 成立,则 $i\in A$ 时

$$p_{ij}^{(n)}=\sum_{k} p_{ik}p_{kj}^{(n-1)}=\sum_{k\in A} p_{ik}p_{kj}^{(n-1)},$$

$$\sum_{j\in A} p_{ij}^{(n)}=\sum_{j\in A}\sum_{k\in A} p_{ik}p_{kj}^{(n-1)}=\sum_{k\in A} p_{ik}\sum_{j\in A} p_{kj}^{(n-1)}=\sum_{k\in A} p_{ik}=1.$$

由(2)式可知,从闭集内的任何一个状态出发,不可能到达闭集外的状态,这就是闭集这个名称的由来.我们应注意,这时$(p_{ij})_{i,j\in A}$本身就是一个随机矩阵.换句话说,当我们只限于在闭集 A 中进行讨论时,可以把闭集 A 作为状态空间对待.

定义 若 $p_{ii}=1$,则状态 i 称为吸收状态.

直观上看,系统一旦进入吸收状态,就一直停留在那里,再也不离开了,也就是被吸收住了.吸收状态必定是本质状态,且本身构成一个类:$C(i)=\{i\}$.自然,$\{i\}$是一个极小闭集.反过来,很容易看出,若一个状态 i 本身构成一个本质类,那么$\{i\}$为闭集,因此 $p_{ii}=1$,i 为吸收状态.因此也可以说,吸收状态就是自身构成一个极小闭集的状态.

容易看出,若状态 i 属于闭集A,则 A 也包含$C(i)$.换句话说,闭集一定是若干个类的并集.下面两个定理使我们能较清楚地了解闭集的结构.

定理2.3 一个状态集合 A 是极小闭集的充要条件为它是一个本质类.

证 先证充分性.设 A 为一本质类.我们已经指出,$i\in A$,但 $j\notin A$ 时,$i\not\to j$.因此 $p_{ij}=0$.所以$i\in A$时

$$\sum_{j\in A} p_{ij}=\sum_{j\in E} p_{ij}=1,$$

即 A 为闭集.若 $B\subset A$,B 也为闭集,但 $B\neq A$,取 $i\in B$,$j\in A\setminus B$.因为 B 是闭集,由(2)式,对一切 $n\geqslant 1$,$p_{ij}^{(n)}=0$,即 $i\nrightarrow j$.但 i,j 属于同一个类必定相通,得出矛盾.所以 A 是极小闭集.

再证必要性.设 A 为极小闭集.任取一个状态 $i\in A$,令 $D(i)=\{j:i\rightarrow j\}$,即从 i 出发可到达的状态全体.显然 $D(i)\subset A$.下证 $D(i)$是闭集.事实上,如果有 $j\in D(i)$,$k\notin D(i)$,但 $p_{jk}>0$,则 $j\rightarrow k$,而 $i\rightarrow j$,故 $i\rightarrow k$,$k\in D(i)$,得出矛盾.$D(i)$显然不是空集(若对一切 $j\neq i$,$i\nrightarrow j$,则 i 为吸收状态,$A=\{i\}=D(i)$),又由于 A 是极小闭集,因此只能 $D(i)=A$.这表明从 A 中任意一个状态出发可到达 A 的任何一个状态,即 A 中任意两个状态都相通,而从 A 中的状态出发只能到达 A 中的状态,因此 A 是一个本质类. □

定理 2.4 有限个非本质类的并集不是闭集.

证 用反证法.设 $C_1,C_2,\cdots,C_n$ 为 n 个非本质类,但 $C=\bigcup_{v=1}^{n}C_v$ 为闭集.

任取 $i_1\in C_{m_1}(m_1=1)$,因为 i_1 非本质,故必有 $i_2\in C$,使 $i_1\rightarrow i_2$,但 $i_2\nrightarrow i_1$.设 $i_2\in C_{m_2}$,由于同一个类中的状态是相通的,因此 $m_2\neq m_1$.所以 $C_{m_1}\rightarrow C_{m_2}$,$C_{m_2}\nrightarrow C_{m_1}$.

对 i_2 也有 $i_3\in C_{m_3}$,使 $i_2\rightarrow i_3$,$i_3\nrightarrow i_2$.同样地有 $C_{m_1}\rightarrow C_{m_2}\rightarrow C_{m_3}$,但 $C_{m_3}\nrightarrow C_{m_2}$,且 $C_{m_3}\nrightarrow C_{m_1}$(如 $C_{m_3}\rightarrow C_{m_1}$,就有 $C_{m_3}\rightarrow C_{m_1}\rightarrow C_{m_2}$与 $C_{m_3}\nrightarrow C_{m_2}$矛盾了),所以 m_3 与 m_1,m_2 也都不相同.按此方法,对任意 k 可取到 C_{m_k},使

$$C_{m_1}\rightarrow C_{m_2}\rightarrow\cdots\rightarrow C_{m_k},$$

$$C_{m_k}\nrightarrow C_{m_{k-1}},\cdots,C_{m_k}\nrightarrow C_{m_1},$$

从而 $m_1,m_2,\cdots,m_k$ 都互不相同.但 $m_1,\cdots,m_k$ 又只能在有限个数 $1,2,\cdots,n$ 中选取,显然矛盾.所以 C 不是闭集. □

定义 整个状态空间显然是一个闭集,如果它又是极小闭集,那么马尔可夫链称为不可约的.

由定理 2.3,不可约链的状态空间是一个本质类.当把马尔可夫链的状态转移局限于一个本质类时,我们就得到一个不可约链.因此从讨论不可约链开始往往可以简化所讨论的问题.

例 2.1 设马尔可夫链的状态空间为$\{0,1,2,3,4\}$,转移概率矩阵为

$$\mathbf{P}=\begin{pmatrix}1/3 & 2/3 & 0 & 0 & 0\\ 0 & 1/2 & 1/2 & 0 & 0\\ 3/4 & 1/4 & 0 & 0 & 0\\ 1/3 & 0 & 0 & 1/3 & 1/3\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}.$$

易见“4”为吸收状态,自身成一个本质类$\{4\}$.不难看出,$0\rightarrow 1$,$1\rightarrow 2$,$2\rightarrow 0$,且$\{0,1,2\}$为闭集.因此$\{0,1,2\}$为本质类.又 $3\rightarrow 1$,$3\rightarrow 4$,但 $1\nrightarrow 3$,$4\nrightarrow 3$,所以$\{3\}$为非本质类. □

从上面的讨论及例 2.1 可看出,一个马尔可夫链的状态的分类只取决于转移概率矩阵中哪些元素是零,哪些元素是正的.但正元素的数值有多大对状态的分类并没有影响.

如果我们把一个有限马尔可夫链的状态编号作适当调整,把非本质状态放在前面,把同一个类的状态放在一起,那么转移概率矩阵有下列分块矩阵的形式:

$$\mathbf{P}=\begin{pmatrix}\mathbf{Q} & \boldsymbol{S}_1 & \boldsymbol{S}_2 & \cdots\\ \boldsymbol{O} & \mathbf{P}_1 & \boldsymbol{O} & \cdots\\ \boldsymbol{O} & \boldsymbol{O} & \mathbf{P}_2 & \cdots\\ \cdots & \cdots & \cdots & \cdots\end{pmatrix},$$

其中 $\mathbf{Q}=(p_{ij})_{i,j\in T}$,$\mathbf{P}_1=(p_{ij})_{i,j\in R_1}$,$\mathbf{P}_2=(p_{ij})_{i,j\in R_2}$,$\cdots$,$T$ 为非本质状态全体,$R_1,R_2,\cdots$ 为本质类(由定理 2.4,有限链必定有本质类).$\mathbf{P}_1,\mathbf{P}_2,\cdots$ 各自都是随机矩阵.分块矩阵中的 $\boldsymbol{O}$ 代表零矩阵,矩阵 $\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots$ 一般不是零矩阵,因为从非本质状态可到达本质状态.无限状态的马尔可夫链的转移概率矩阵也可类似地写成适当的分块矩阵形式.

例 2.2 设马尔可夫链的转移概率矩阵为

$$\mathbf{P}=\begin{pmatrix}1/2 & 1/2 & 0 & 0 & \cdots\\ 0 & 1/2 & 1/2 & 0 & \cdots\\ 0 & 0 & 1/2 & 1/2 & \cdots\\ \cdots & \cdots & \cdots & \cdots & \cdots\end{pmatrix},$$

$$p_{ii}=p_{i,i+1}=1/2,\qquad i\geqslant 0.$$

这是一个极端的例子,我们用它来说明无限多个非本质类的并集确实可以是闭集.因为这里每一个状态构成一个非本质类,它们的并集即状态空间,自然是闭集.这例子还表明一个闭集可以不包含极小闭集,因为每一个形为 $\{i,i+1,i+2,\cdots\}$ 的集合都是闭集. □

定义 若对状态 i,正整数集 $\{n\geqslant 1: p_{ii}^{(n)}>0\}$ 非空,它的最大公因子 d_i 称为状态的周期.若 $d_i>1$,状态 i 称为周期的;若 $d_i=1$,状态 i 称为非周期的.

若 $\{n\geqslant 1: p_{ij}^{(n)}>0\}$ 为空集,即 $i\nrightarrow i$,对这样的状态我们不讨论它的周期.这时状态 i 必定是非本质状态,且自身构成一个类.实际上,我们一般只对本质状态注意它的周期.

定理 2.5 若 $i\leftrightarrow j$,则 $d_i=d_j$.

证 设 $l\geqslant 1,n\geqslant 1$,使 $p_{ij}^{(n)}>0,p_{ji}^{(l)}>0$,则

$$p_{ii}^{(l+n)}\geqslant p_{ij}^{(n)}p_{ji}^{(l)}>0,\quad p_{jj}^{(l+n)}\geqslant p_{ji}^{(l)}p_{ij}^{(n)}>0,$$

$l+n$ 同时能被 d_i 及 d_j 整除,若 $p_{ii}^{(m)}>0$,则

$$p_{jj}^{(l+m+n)}\geqslant p_{ji}^{(l)}p_{ii}^{(m)}p_{ij}^{(n)}>0,$$

$l+m+n$ 也被 d_j 整除,因此 m 要被 d_j 整除,从而 d_j 整除 $\{m\geqslant 1: p_{ii}^{(m)}>0\}$ 的最大公因子 d_i.对调 i 与 j 的位置,d_i 也整除 d_j,所以 $d_i=d_j$. □

定义 一个本质类的周期定义为该类中任一状态的周期.称周期大于 1 的本质类为周期类,称周期为 1 的本质类为非周期类.

定理 2.6 设 C 为本质类,周期为 d,$i,j\in C$,$p_{ij}^{(n_1)}>0,p_{ij}^{(n_2)}>0$,则 n_2-n_1 被 d 整除.

证 因为 $j\rightarrow i$,有 $n\geqslant 1$ 使 $p_{ji}^{(n)}>0$,从而

$$p_{ii}^{(n_1+n)}\geqslant p_{ij}^{(n_1)}p_{ji}^{(n)}>0,\quad p_{ii}^{(n_2+n)}\geqslant p_{ij}^{(n_2)}p_{ji}^{(n)}>0,$$

n_1+n,n_2+n 均被 d 整除,故 n_2-n_1 被 d 整除. □

根据定理 2.6,在周期本质类中,从一个状态进入另一个状态,必须是周期地发生的,所以可以把一个周期为 d 的周期本质类 C 划分为 d 个互不相交的子类.任取 C 中一个状态 i,令 $j\in C_s(i)$当且仅当 $p_{ij}^{(n)}>0$ 时必须有 $n=kd+s,k\geqslant 0,0\leqslant s\leqslant d-1$,那么

$$C(=C(i))=C_0(i)\cup C_1(i)\cup\cdots\cup C_{d-1}(i).$$

当然也可取 C 中另一个状态 j 来划分子类.但是如果 $j\in C_r(i)$,则 $C_s(j)=C_{r+s}(i)$(若 $n=kd+s,k\geqslant 1,0\leqslant s\leqslant d-1$,定义 $C_n(i)=C_s(i)$).也就是说,$\{C_0(i),\cdots,C_{d-1}(i)\}$与$\{C_0(j),\cdots,C_{d-1}(j)\}$两个集合完全一样,只是排列顺序不同而已.因此可以把这 d 个子类记为 $C_0,C_1,\cdots,C_{d-1}$(若 $n=kd+s,k\geqslant 1,0\leqslant s\leqslant d-1$,定义 $C_n=C_s$).这时对一切 $n\geqslant 1$

$$\sum_{j\in C_{r+n}}p_{ij}^{(n)}=1,\quad i\in C_r.$$

对 $i\in C_r$, $\sum_{j\in C_{r+1}}p_{ij}=1$ 的直观意义是很清楚的:从状态 $i\in C$ 出发走一步只能进入 C_{r+1}中的状态,再走一步只能进入 C_{r+2}中的状态,等等.换句话说,系统的状态在周期本质类中的转移是依次从 $C_0,C_1,\cdots,C_{d-1}$中的一个子类进到下一个子类,一直这样地循环进行下去.如果我们把状态转移局限在这个周期本质类中,并把原来的 d 步转移概率矩阵作为一步转移概率矩阵,也就是把原来的 d 步看作一步,把时间单位放大 d 倍,那么我们得到一个新的马尔可夫链.在新的链中状态就都是非周期的了.容易看出,子类 $C_0,C_1,\cdots,C_{d-1}$在新的链中都成为闭集了.这是一种把讨论周期本质状态化为讨论非周期状态的有用的方法.

例 2.3 设马尔可夫链的状态空间为$\{0,1,2,3,4,5\}$,转移概率矩阵为

$$\mathbf{P}=\begin{pmatrix}0&1/4&1/4&0&1/2&0\\0&1/3&0&2/3&0&0\\0&0&0&1/2&0&1/2\\0&0&1/3&0&2/3&0\\0&0&0&1/4&0&3/4\\0&0&4/5&0&1/5&0\end{pmatrix}.$$

先将状态分类:$\{2,3,4,5\}$为本质类,$\{0\}$及$\{1\}$为非本质类.对一切 $n\geqslant 1,p_{00}^{(n)}=0$,因此我们不讨论状态 0 的周期.$p_{11}^{(1)}=1/3>0$,由此即知$\{n\geqslant 1:p_{ii}^{(n)}>0\}$的最大公因子只能为 1,即 $d_1=1$,状态 1 为非周期状态.在本质类$\{2,3,4,5\}$中走一步从偶数状态只能到达奇数状态,从奇数状态只能到达偶数状态.因此周期为 2,且$\{2,4\}$及$\{3,5\}$为两个子类. □

下面我们先给出一个初等数论的结果.

引理 2.1 设 $m\geqslant 2$,正整数 $s_1,s_2,\cdots,s_m$ 的最大公因子为 d,则存在正整数 N,使得 $n>N$ 时,必有非负整数 $c_1,c_2,\cdots,c_m$ 使 $nd=\sum_{i=1}^m c_is_i$.

证 由于 d 是 $s_1,\cdots,s_m$ 的最大公因子,必定存在整数 $n_1,\cdots,n_m$ 使 $\sum_{i=1}^m n_is_i=d$.($m=2$ 时由求最大公因子的辗转相除法即可得这结论,对一般的 m 再用归纳法证明.)令

$$q=\sum_{n_i\geqslant 0}n_is_i,\qquad r=\sum_{n_i<0}(-n_i)s_i,$$

则 $q=r+d$.若 $r=0$,则 $q=d$.这时必须有一个 $n_i=1$,其余的 n_j 全为零,$s_i=d$,从而引理的结论显然成立.若 $r\neq 0$,取 $N=\left(\frac{r}{d}\right)^2$,则 $n>N$ 时,把 n 写成

$$n=k\left(\frac{r}{d}\right)+l,\qquad 0\leqslant l\leqslant\frac{r}{d}-1,$$

必定有 $k\geqslant\frac{r}{d}>l$,$k-l>0$,从而

$$nd=(k-l)r+l(r+d)=(k-l)r+lq=\sum_{n_i\geqslant 0}(ln_i)s_i+\sum_{n_i<0}(k-l)(-n_i)s_i.\qquad\square$$

定理 2.7 设状态 i 的周期为 d,则存在正整数 N,使得 $n\geqslant N$ 时,$p_{ii}^{(nd)}>0$.

证 这时存在 $s_1,s_2,\cdots,s_m$,使得它们的最大公因子为 d,且 $p_{ii}^{(s_j)}>0,j=1,\cdots,m$.由引理 2.1,存在正整数 N,使得 $n\geqslant N$ 时,必有非负整数 $c_1,c_2,\cdots,c_m$ 使 $nd=\sum_{i=1}^m c_is_i$,从而

$$p_{ii}^{(nd)}\geqslant(p_{ii}^{(s_1)})^{c_1}\cdots(p_{ii}^{(s_m)})^{c_m}>0.\qquad\square$$

例 2.4 设马尔可夫链 $X=\{X_n,n\geqslant 0\}$ 为非周期不可约的,马尔可夫链 $Y=\{Y_n,n\geqslant 0\}$ 与 X 相互独立,且有相同的转移概率矩阵 (p_{ij}).令

$$Z_n=(X_n,Y_n),\qquad n\geqslant 0,$$

则 $Z=\{Z_n,n\geqslant 0\}$ 是状态空间为 $E\times E$ 的非周期不可约马尔可夫链.事实上,对任意的 $n\geqslant 1$, $i_0,j_0,\cdots,i_{n+1},j_{n+1}\in E$

$$\begin{aligned}
&\mathrm{P}(Z_{n+1}=(i_{n+1},j_{n+1})\mid Z_0=(i_0,j_0),\cdots,Z_n=(i_n,j_n))\\
&=\frac{\mathrm{P}(\bigcap_{k=0}^{n+1}\{X_k=i_k,Y_k=j_k\})}{\mathrm{P}(\bigcap_{k=0}^{n}\{X_k=i_k,Y_k=j_k\})}\\
&=\frac{\mathrm{P}(\bigcap_{k=0}^{n+1}\{X_k=i_k\})\mathrm{P}(\bigcap_{k=0}^{n+1}\{Y_k=j_k\})}{\mathrm{P}(\bigcap_{k=0}^{n}\{X_k=i_k\})\mathrm{P}(\bigcap_{k=0}^{n}\{Y_k=j_k\})}\\
&=\mathrm{P}(X_{n+1}=i_{n+1}\mid X_n=i_n,\cdots,X_0=i_0)\\
&\qquad\cdot\mathrm{P}(Y_{n+1}=j_{n+1}\mid Y_n=j_n,\cdots,Y_0=j_0)\\
&=\mathrm{P}(X_{n+1}=i_{n+1}\mid X_n=i_n)\mathrm{P}(Y_{n+1}=j_{n+1}\mid Y_n=j_n).
\end{aligned}\tag{3}$$

类似地有

$$\begin{aligned}
&\mathrm{P}(Z_{n+1}=(i_{n+1},j_{n+1})\mid Z_n=(i_n,j_n))\\
&\qquad=\mathrm{P}(X_{n+1}=i_{n+1}\mid X_n=i_n)\mathrm{P}(Y_{n+1}=j_{n+1}\mid Y_n=j_n).
\end{aligned}\tag{4}$$

由(3)及(4)即得 Z 的马尔可夫性.由(3)还可看出,Z 的一步转移概率为

$$p_{(i,j)(k,l)}=p_{ik}p_{jl}.\tag{5}$$

用同样的方法也可知道 Z 的 n 步转移概率为

$$p_{(i,j)(k,l)}^{(n)}=p_{ik}^{(n)}p_{jl}^{(n)}.\tag{6}$$

对任意的 $i,k\in E$,存在 $m_{ik}\geqslant 1$ 使 $p_{ik}^{(m_{ik})}>0$ 及 $m_k\geqslant 1$ 使 $n\geqslant m_k$ 时,$p_{kk}^{(n)}>0$,从而 $n\geqslant N_{ik}=$

$m_{ik}+m_k$ 时

$$p_{ik}^{(n)} \geq p_{ik}^{(m_{ik})} p_{kk}^{(n-m_{ik})} > 0.$$

因此,对任意的$(i,j)(k,l)\in E\times E$,$n\geq N=\max(N_{ik},N_{jl})$时由(6)可知 $p_{(i,j)(k,l)}^{(n)}>0$.由此可得,马尔可夫链 Z 也是非周期不可约的. □

习　题

2-1. 讨论具有下列转移概率矩阵的马尔可夫链的状态的分类(区别本质与非本质)与周期:

(1) $\begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$;　　(2) $\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$;

(3) $\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$;　　(4) $\begin{pmatrix} 1/3 & 1/3 & 0 & 0 & 1/3 \\ 1/2 & 1/4 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$;

(5) $\begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 & \cdots \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & \cdots \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 & \cdots \\ 1/4 & 0 & 0 & 1/4 & 1/2 & 0 & \cdots \\ 1/4 & 0 & 0 & 0 & 1/4 & 1/2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$;

(6) $\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & \cdots \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & \cdots \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & \cdots \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$.

2-2. 设每次从数字 $1,2,\cdots,a$ 中任取一数(可重复取).以 X_n 记前 n 次所取数中的最大数,则$\{X_n,n\geq1\}$为马尔可夫链,讨论它的状态的分类与周期.

2-3. 逐个随机地把球放入 a 个盒子中(可重复放入).以 X_n 记放了 n 个球之后的空盒数,则$\{X_n,n\geq0\}$为马尔可夫链,讨论它的状态的分类.

2-4. 设一个细菌群体包含有 a 个细菌,但细菌有两种类型,每个细菌分裂为类型相同的两个细菌,然后从 $2a$ 细菌中随机地挑选 a 个构成下一代.每代的挑选是相互独立的.以 X_n 记第 n 代中每一种类型的细菌的个数,则$\{X_n,n\geq0\}$为马尔可夫链,讨论它的状态分类.

2-5. 把数字 $1,2,\cdots,a$ 作随机的排列:

$$b_1,b_2,\cdots,b_a.$$

令 $X_0=1$. 对 $n\geqslant 0$,定义

$$X_{n+1}=\min\{i:b_i>b_{X_n},1\leqslant i\leqslant a\},$$

若 $\{i:b_i>b_{X_n},1\leqslant i\leqslant a\}$ 为空集,定义 $X_{n+1}=a+1$,并规定 $b_{a+1}=a+1$,则 $\{X_n,n\geqslant 0\}$ 为马尔可夫链,讨论它的状态分类.

2-6. 证明:$i\to j$ 的充要条件为存在 $n\geqslant 1$ 及 $i_1,\cdots,i_{n-1}\in E$,使得

$$p_{ii_1}p_{i_1i_2}\cdots p_{i_{n-1}j}>0,$$

且 $i,i_1,\cdots,i_{n-1}$ 互不相同.

2-7. 设有限马尔可夫链有 a 个状态. 证明:若 $i\to j$,则必有 n,使 $1\leqslant n\leqslant a$,且 $p_{ij}^{(n)}>0$.

2-8. 设 A 为一状态集合使得集合

$$\hat{A}=\{i\in E:\ 对一切\ j\in A,i\not\to j\}$$

非空,则 $\hat{A}$ 及 $\hat{A}\setminus A$ 均为闭集.

2-9. 设有 $n\geqslant 1$ 使得对一切 i,$p_{ij}^{(n)}>0$. 证明:j 是非周期状态.

2-10. 设 $\mathbf{P}$ 为有限马尔可夫链的转移概率矩阵. 证明:$\mathbf{P}$ 的特征值的模不大于 1,且 1 为 $\mathbf{P}$ 的特征值.

若 $\mathbf{P}$ 又为非周期、不可约的,证明:若 λ 为 $\mathbf{P}$ 的特征值,且 $|\lambda|=1$,则 $\lambda=1$.

§2.3 常 返 性

我们继续对一个固定的马尔可夫链 $X=\{X_n,n\geqslant 0\}$ 讨论,其转移概率矩阵为 (p_{ij}). 对任意一个状态 j,定义

$$r_j=\inf\{n\geqslant 1:X_n=j\}$$

为 $\{X_n,n\geqslant 1\}$ 首次进入状态 j 的时刻. 如果 $\{n\geqslant 1:X_n=j\}$ 为空集,按定义 $r_j=+\infty$. r_j 是停时,因为对任意 n,事件

$$\{r_j=n\}=\{X_1\neq j,\cdots,X_{n-1}\neq j,X_n=j\}$$

由 $X_1,\cdots,X_n$ 的取值所决定. 令

$$\begin{aligned}f_{ij}^{(n)}&=\mathsf{P}_i(r_j=n)=\mathsf{P}_i(X_n=j,X_v\neq j,0<v<n)\\&=\sum_{i_v\neq j,0<v<n}p_{ii_1}p_{i_1i_2}\cdots p_{i_{n-1}j}.\end{aligned}$$

我们以 $\mathsf{P}_i(A)$ 记在条件 $X_0=i$ 之下事件 A 的条件概率 $\mathsf{P}(A\mid X_0=i)$. $f_{ij}^{(n)}$ 是从状态 i 出发经过 n 步首次到达状态 j 的概率. 因此 $f_{ij}=\sum\limits_{n=1}^{\infty}f_{ij}^{(n)}$ 是从状态 i 出发迟早要到达状态 j 的概率,或从状态 i 出发至少到达状态 j 一次的概率:

$$f_{ij}=\mathsf{P}_i(r_j<\infty)=\mathsf{P}_i\Big(\bigcup_{n=1}^{\infty}\{X_n=j\}\Big).$$

事实上，只需把 $\bigcup_{n=1}^{\infty}\{X_n=j\}$ 写成互不相容的事件的并：

$$\bigcup_{n=1}^{\infty}\{X_n=j\}=\bigcup_{n=1}^{\infty}\{X_n=j,X_\upsilon\neq j,0<\upsilon<n\}=\bigcup_{n=1}^{\infty}\{\tau_j=n\}=\{\tau_j<\infty\}.$$

若 j 是吸收状态，f_{ij} 也称为吸收概率，即从状态 i 出发，被吸收状态 j 吸收的概率. 这时对一切 $n\geqslant 1$，我们有 $\{X_n=j\}\subset\{X_{n+1}=j\}$（一旦进入吸收状态 j，就永远停留在 j 上），因此吸收概率

$$f_{ij}=\lim_{n\to\infty}\mathrm{P}_i(\bigcup_{\upsilon=1}^{n}\{X_\upsilon=j\})=\lim_{n\to\infty}\mathrm{P}_i(X_n=j)=\lim_{n\to\infty}p_{ij}^{(n)}.$$

易见，对一切 $i,j\in E$ 及 $n\geqslant 1$

$$0\leqslant f_{ij}^{(n)}\leqslant p_{ij}^{(n)}\leqslant f_{ij}\leqslant 1.$$

定理 3.1 $i\to j$ 的充要条件是 $f_{ij}>0$.

证 若 $i\to j$，则有 $n\geqslant 1$，使 $p_{ij}^{(n)}>0$，从而 $f_{ij}\geqslant p_{ij}^{(n)}>0$. 反之，若 $f_{ij}>0$，则有 $n\geqslant 1$ 使 $f_{ij}^{(n)}>0$，从而 $p_{ij}^{(n)}\geqslant f_{ij}^{(n)}>0$. 因此 $i\to j$. □

定义 若 $f_{ii}=1$，称状态 i 为常返的；若 $f_{ii}<1$，称状态 i 为非常返的或滑过的.

下面我们将解释这些名称的由来. 对于常返状态 i，$\mathrm{P}(\tau_i<\infty\mid X_0=i)=f_{ii}=1$，这表明从状态 i 出发必定要返回到状态 i，且 $\{f_{ii}^{(n)},n\geqslant 1\}$ 就是返回时间的概率分布（τ_i 在 $X_0=i$ 的条件下的条件分布）. 我们要进一步证明，从常返状态 i 出发必定要无穷次地返回到状态 i，这就是常返的含义. 为此我们定义 g_{ij} 为从状态 i 出发无穷多次到达状态 j 的概率：

$$g_{ij}=\mathrm{P}_i(\text{有无穷多个 } n\geqslant 1 \text{ 使 } X_n=j)=\mathrm{P}_i(\bigcap_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\{X_n=j\}),$$

事实上，若有无穷多个 $n\geqslant 1$，使 $X_n=j$，则对一切 $k\geqslant 1$，事件 $\bigcup_{n=k}^{\infty}\{X_n=j\}$ 都发生，所以事件 $\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\{X_n=j\}$ 发生；反之，若事件 $\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\{X_n=j\}$ 发生，则对任一 $k\geqslant 1$，必有 $n\geqslant k$，使 $X_n=j$，所以有无穷多个 $n\geqslant 1$，使 $X_n=j$. 显然，我们有

$$0\leqslant g_{ij}\leqslant f_{ij}.$$

定理 3.2 对一切 i,j，

$$g_{ii}=\lim_{n\to\infty}(f_{ii})^n,\qquad g_{ij}=f_{ij}g_{jj}.\tag{1}$$

证 以 $g_{ij}(m)$ 记从状态 i 出发至少到达状态 j m 次的概率：

$$g_{ij}(m)=\mathrm{P}_i(\text{至少有 } m \text{ 个 } n\geqslant 1 \text{ 使 } X_n=j).$$

显然 $g_{ij}(1)=f_{ij}$. 因为事件 $\{$至少有 m 个 $n\geqslant 1$ 使 $X_n=j\}$ 是随 m 增加而单调下降的，且

$$\bigcap_{m=1}^{\infty}\{\text{至少有 } m \text{ 个 } n\geqslant 1 \text{ 使 } X_n=j\}=\{\text{有无穷多个 } n\geqslant 1 \text{ 使 } X_n=j\}.$$

因此，$g_{ij}(m)\downarrow g_{ij}$，$m\to\infty$. 对 $m\geqslant 0$，

$$
\begin{aligned}
g_{ij}(m+1) &= \mathsf{P}_i(\text{至少有 } m+1 \text{ 个 } n\geqslant 1 \text{ 使 } X_n=j) \\
&= \sum_{k=1}^{\infty} \mathsf{P}_i(r_j=k;\text{至少有 } m+1 \text{ 个 } n\geqslant 1 \text{ 使 } X_n=j) \\
&= \sum_{k=1}^{\infty} \mathsf{P}_i(X_k=j,X_v\neq j,0<v<k;\text{至少有 } m \text{ 个 } n\geqslant k+1 \text{ 使 } X_n=j) \\
&= \sum_{k=1}^{\infty} \mathsf{P}_i(X_k=j,X_v\neq j,0<v<k) \\
&\quad \cdot \mathsf{P}(\text{至少有 } m \text{ 个 } n\geqslant k+1 \text{ 使 } X_n=j\mid X_k=j) \\
&= \sum_{k=1}^{\infty} f_{ij}^{(k)}\mathsf{P}_j(\text{至少有 } m \text{ 个 } n\geqslant 1 \text{ 使 } X_n=j) \\
&= \sum_{k=1}^{\infty} f_{ij}^{(k)} g_{jj}(m) = f_{ij}g_{jj}(m),
\end{aligned}
$$

由此特别得到

$$
g_{ii}(m+1)=f_{ii}g_{ii}(m)=\cdots=(f_{ii})^m g_{ii}(1)=(f_{ii})^{m+1}. \tag{2}
$$

从而一般地有

$$
g_{ij}(m+1)=f_{ij}(f_{jj})^m, \tag{3}
$$

在(2)及(3)式中令 $m\to\infty$ 即得(1)式. □

由(1)式，i 为常返状态时，$g_{ii}=1$，即从状态 i 出发要无限多次返回到状态 i，这就是我们早已指出的常返的意义. i 为非常返状态时，$g_{ii}=0$，即从状态 i 出发至多返回状态 i 有限多次，以后就再也不回到状态 i 了(但必须注意返回的次数是随机变量)，这就是“滑过”的意义. 一般地，若 j 为常返状态，则对一切 i，$g_{ij}=f_{ij}$；若 j 为非常返状态，则对一切 i，$g_{ij}=0$，即从状态 i 出发至多到达状态 j 有限多次.

定理 3.2 的证明方法同样是我们早已指出的研究马尔可夫链的典型方法，但这里事件分解的方法有所不同. 我们把概率 $g_{jj}(m+1)$ 按首次进入状态 j 的时刻 r_j 的取值进行分解，然后再利用马尔可夫性及齐次性. 为了强调这一事件分解的方法，通常把它称为首次进入法.

定理 3.3 设 i 为常返状态，$i\to j$，则

$$
g_{ji}=f_{ji}=1.
$$

证 对任意 $m\geqslant 1$ 及 $l\in E$

$$
\begin{aligned}
g_{il} &= \mathsf{P}_i(\text{有无穷多个 } n\geqslant 1 \text{ 使 } X_n=l) \\
&= \sum_k \mathsf{P}_i(X_m=k;\text{有无穷多个 } n\geqslant 1 \text{ 使 } X_n=l) \\
&= \sum_k \mathsf{P}_i(X_m=k;\text{有无穷多个 } n\geqslant m+1 \text{ 使 } X_n=l) \\
&= \sum_k \mathsf{P}_i(X_m=k)\mathsf{P}(\text{有无穷多个 } n\geqslant m+1 \text{ 使 } X_n=l\mid X_m=k) \\
&= \sum_k p_{ik}^{(m)}\mathsf{P}_k(\text{有无穷多个 } n\geqslant 1 \text{ 使 } X_n=l) = \sum_k p_{ik}^{(m)} g_{kl}.
\end{aligned}
$$

因为 i 是常返的,由定理 3.2

$$0=1-g_{ii}=\sum_{k}p_{ik}^{(m)}(1-g_{ki}).$$

从而对一切 $m\geqslant1$ 及 $k\in E$, $p_{ik}^{(m)}(1-g_{ki})=0$. 若 $i\to j$, 必有 $m\geqslant1$ 使 $p_{ij}^{(m)}>0$, 这时必须 $g_{ji}=1$. 但是 $f_{ji}\geqslant g_{ji}$, 因此 $f_{ji}=1$. □

推论 常返状态是本质状态.

下面我们需要一个十分有用的数学分析结果,我们把它写成一个引理.

引理 3.1 设 $\{a_n,n\geqslant0\}$ 为一个不全为零的非负数列,且满足条件

$$\lim_{n\to\infty}\frac{a_n}{\sum_{\upsilon=0}^{n}a_\upsilon}=0,\tag{4}$$

$\{b_n,n\geqslant0\}$ 为一个收敛数列,则

$$\lim_{n\to\infty}\frac{\sum_{\upsilon=0}^{n}a_\upsilon b_{n-\upsilon}}{\sum_{\upsilon=0}^{n}a_\upsilon}=\lim_{n\to\infty}b_n.\tag{5}$$

证 设 $b=\lim b_n$. 对任给 $\varepsilon>0$, 存在 N, 使 $n\geqslant N$ 时 $|b_n-b|\leqslant\varepsilon$. 又设 $B=\sup_{n\geqslant0}|b_n|$, 则 $n\geqslant N$ 时

$$\left|\sum_{\upsilon=0}^{n}a_\upsilon(b_{n-\upsilon}-b)\right|\leqslant\varepsilon\sum_{\upsilon=0}^{n-N}a_\upsilon+2B\sum_{\upsilon=N+1}^{n}a_\upsilon.$$

对固定的 N, 由条件(4), $n\to\infty$ 时,

$$\frac{\sum_{\upsilon=n-N+1}^{n}a_\upsilon}{\sum_{\upsilon=0}^{n}a_\upsilon}\to0,\qquad\frac{\sum_{\upsilon=0}^{n-N}a_\upsilon}{\sum_{\upsilon=0}^{n}a_\upsilon}\to1,$$

由此得到

$$\varlimsup_{n\to\infty}\left|\frac{\sum_{\upsilon=0}^{n}a_\upsilon b_{n-\upsilon}}{\sum_{\upsilon=0}^{n}a_\upsilon}-b\right|\leqslant\varepsilon,$$

再令 $\varepsilon\to0$ 即得(5). □

定理 3.4 对任意的 $n\geqslant1$ 及 $i,j\in E$

$$p_{ij}^{(n)}=\sum_{\upsilon=1}^{n}f_{ij}^{(\upsilon)}p_{jj}^{(n-\upsilon)}.\tag{6}$$

证 用首次进入法,并注意到 $X_n=j$ 时 $r_j\leqslant n$,

$$
\begin{aligned}
p_{ij}^{(n)} = \mathrm{P}_i(X_n = j) &= \sum_{\upsilon=1}^{n} \mathrm{P}_i(r_j = \upsilon, X_n = j) \\
&= \sum_{\upsilon=1}^{n} \mathrm{P}_i(X_\upsilon = j, X_k \neq j, 0 < k < \upsilon, X_n = j) \\
&= \sum_{\upsilon=1}^{n} \mathrm{P}_i(X_\upsilon = j, X_k \neq j, 0 < k < \upsilon) \mathrm{P}(X_n = j \mid X_\upsilon = j) \\
&= \sum_{\upsilon=1}^{n} f_{ij}^{(\upsilon)} p_{jj}^{(n-\upsilon)}. \quad \square
\end{aligned}
$$

定理 3.5 对任意的 $i, j \in E$,

$$
\lim_{N \to \infty} \frac{\sum_{n=1}^{N} p_{ij}^{(n)}}{\sum_{n=0}^{N} p_{jj}^{(n)}} = f_{ij}. \tag{7}
$$

证 把(6)式从 $n=1$ 加到 $n=N$,并交换和号:

$$
\begin{aligned}
\sum_{n=1}^{N} p_{ij}^{(n)} &= \sum_{n=1}^{N} \sum_{\upsilon=1}^{n} f_{ij}^{(\upsilon)} p_{jj}^{(n-\upsilon)} = \sum_{n=1}^{N} \sum_{\upsilon=0}^{n-1} f_{ij}^{(n-\upsilon)} p_{jj}^{(\upsilon)} \\
&= \sum_{\upsilon=0}^{N-1} \sum_{n=\upsilon+1}^{N} f_{ij}^{(n-\upsilon)} p_{jj}^{(\upsilon)} = \sum_{\upsilon=0}^{N-1} p_{jj}^{(\upsilon)} \sum_{n=1}^{N-\upsilon} f_{ij}^{(n)}.
\end{aligned} \tag{8}
$$

我们要利用引理 3.1,取

$$
a_n = p_{jj}^{(n)}, n \geqslant 0, \quad b_n = \sum_{\upsilon=1}^{n} f_{ij}^{(\upsilon)}, n \geqslant 1, b_0 = 0,
$$

现在验证条件(4)成立.在 $\sum_n p_{jj}^{(n)} < \infty$ 时,$\lim p_{jj}^{(n)} = 0$;在 $\sum_n p_{jj}^{(n)} = \infty$ 时,由于 $p_{jj}^{(n)} \leqslant 1$,仍然有 $p_{ij}^{(n)} \Big/ \sum_{\upsilon=0}^{n} p_{ij}^{(\upsilon)} \to 0$.总之,条件(4)成立.再由(8)式,并用引理 3.1 即得(7)式. $\square$

通常把定理 3.5 及与(7)形式相似的结果称为比极限定理.由这个比极限定理,我们可得到一个常返的判别法.

定理 3.6 (1) 状态 i 为常返的充要条件是

$$
\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty; \tag{9}
$$

(2) 状态 i 为非常返的充要条件是

$$
\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty, \tag{10}
$$

且这时

$$
\sum_{n=0}^{\infty} p_{ii}^{(n)} = \frac{1}{1 - f_{ii}}. \tag{11}
$$

证 由(7)式以及 $p_{ij}^{(0)} = \delta_{ij}$,

$$\lim_{N\to\infty}\frac{\sum_{n=1}^{N}p_{ii}^{(n)}}{\sum_{n=0}^{N}p_{ii}^{(n)}}=f_{ii},\qquad \lim_{N\to\infty}\frac{1}{\sum_{n=0}^{N}p_{ii}^{(n)}}=1-f_{ii},$$

由此,$f_{ii}=1$ 等价于(9)式,$f_{ii}<1$ 等价于(10)式,且(11)式成立. □

推论 1 若 i 为常返状态,$i\to j$,则 j 也为常返状态.

证 i 为常返状态,必为本质的.因此 $i\leftrightarrow j$,有 $m\geqslant1,n\geqslant1$ 使 $p_{ij}^{(n)}>0,p_{ji}^{(m)}>0$.这时

$$p_{jj}^{(m+v+n)}\geqslant p_{ji}^{(m)}p_{ii}^{(v)}p_{ij}^{(n)},$$

$$\sum_{v=0}^{\infty}p_{jj}^{(v)}\geqslant\sum_{v=0}^{\infty}p_{jj}^{(m+v+n)}\geqslant p_{ji}^{(m)}p_{ij}^{(n)}\sum_{v=0}^{\infty}p_{ii}^{(v)}=\infty,$$

由此可知 j 为常返状态. □

由推论 1 可知,同一个类的状态或全为常返的,或全为非常返的.相应地,我们就称这个类为常返类或非常返类.由定理 2.3 我们还知道,若 i,j 属于同一个常返类,则 $g_{ij}=f_{ij}=1$,也就是说,从常返类中的任意一个状态出发必须无穷多次地经过类中任何一个状态.与此相反,从非常返类中的任意一个状态出发,只能有限多次经过类中的任何一个状态.所以,当非常返类只包含有限多个状态时,从这类中任一状态出发,迟早要离开这个类.特别,非常返的本质类必须包含无穷多个状态,因为本质类是闭集,从本质类中的状态出发是不能离开这个类的.由此可得下述推论.

推论 2 有限马尔可夫链的本质状态都是常返状态;从任一非本质状态出发,必定要到达常返状态.不可约有限马尔可夫链的状态空间是一个常返类.

证 由于有限的非常返本质类不存在,所以对有限马尔可夫链,本质状态均为常返的.另一方面,从非本质(也是非常返)状态出发,迟早要离开有限的非常返类,因此必定要到达常返状态.不可约链的状态空间是一个本质类,故为常返类. □

定理 3.7 若 $g_{ij}=0$,则 $\sum_n p_{ij}^{(n)}<\infty$;若 $g_{ij}>0$,则 $\sum_n p_{ij}^{(n)}=\infty$.

证 设 $g_{ij}=0$,由定理 3.2,$f_{ij}g_{jj}=0$.若 $f_{ij}=0$,则对一切 $n\geqslant1$,$p_{ij}^{(n)}=0$,显然有 $\sum_n p_{ij}^{(n)}<\infty$.若 $g_{jj}=0$,则 $\sum_n p_{jj}^{(n)}<\infty$,由定理 3.5,$\sum_n p_{ij}^{(n)}<\infty$.若 $g_{ij}=f_{ij}g_{jj}>0$,则由 $g_{jj}>0$ 知 $g_{jj}=1$,即 j 为常返状态,又由于 $f_{ij}>0$,从而由定理 3.5 知 $\sum_n p_{ij}^{(n)}=\infty$. □

现在我们对级数 $\sum_n p_{ij}^{(n)}$ 的敛散情况已完全清楚:若 j 为非常返状态,则对一切 i,$\sum_n p_{ij}^{(n)}<\infty$;若 j 为常返状态,则 $i\to j$ 时,$\sum_n p_{ij}^{(n)}=\infty$,而 $i\not\to j$ 时,该级数项项为零.

我们再进一步解释级数 $\sum_{n=1}^{\infty}p_{ij}^{(n)}$ 的概率意义.令 N_j 为状态 j 在 $\{X_1,X_2,\cdots,X_n,\cdots\}$ 中出现的次数:

$$N_j=\sum_{n=1}^{\infty}1_{\{X_n=j\}}$$

(1_A 是集合 A 的示性函数).因此前面所定义的 f_{ij} 及 g_{ij} 可写成:

$$f_{ij}=\mathsf{P}_i(N_j\geqslant1),\qquad g_{ij}=\mathsf{P}_i(N_j=\infty).$$

不难看出 N_j 在 $X_0=i$ 的条件下的数学期望为:

$$\mathrm{E}_i[N_j]=\sum_{n=1}^{\infty}\mathrm{E}_i[1_{\{X_n=j\}}]=\sum_{n=1}^{\infty}\mathrm{P}_i(X_n=j)=\sum_{n=1}^{\infty}p_{ij}^{(n)}.$$

当 i 为常返状态时，$g_{ii}=\mathrm{P}_i(N_i=\infty)=1$，自然有 $\sum_{n=1}^{\infty}p_{ii}^{(n)}=\mathrm{E}_i[N_i]=\infty$. 当 i 为非常返状态时，我们虽然知道 $g_{ii}=\mathrm{P}_i(N_i=\infty)=0$，但不能立即断定 N_i 的数学期望有限. 所以还是需要借用比极限定理来得到定理 3.6 中的常返判别法.

例 3.1 对称随机游动 我们考虑(p,q)－随机游动. 已经知道 n 步转移概率为

$$p_{ij}^{(n)}=\begin{cases}\mathrm{C}_n^{(n+j-i)/2}p^{(n+j-i)/2}q^{(n-j+i)/2}, & n+j-i \text{ 为偶数},\\ 0, & n+j-i \text{ 为奇数}.\end{cases}$$

$j>i$ 时，$p_{ij}^{(j-i)}=p^{j-i}>0$，$p_{ji}^{(j-i)}=q^{j-i}>0$，即任意两个状态相通，所以链是不可约的. 易见这是一个周期链. 周期为 2.

利用斯特林(Stirling)公式：$n!\simeq\sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n$，我们有

$$p_{ii}^{(2n)}=\frac{(2n)!}{(n!)^2}(pq)^n\simeq\frac{(4pq)^n}{\sqrt{\pi n}}.$$

若 $p\neq1/2$，则 $4pq<1$，$\sum_n p_{ii}^{(2n)}<\infty$，链是非常返的. 若 $p=1/2$，即对称随机游动，则 $4pq=1$，$\sum_n p_{ii}^{(2n)}=\infty$，从而链是常返的.

现在我们讨论平面上的对称随机游动. 质点的位置是平面上的整数格点(坐标为整数的点). 每个位置有四个相邻的位置，质点各以 1/4 的概率转移到这四个相邻位置中的每一个. 易见平面上的对称游动也是周期为 2 的不可约链.

我们来计算质点经过 $2n$ 步仍回原位置的概率 u_n. 这时质点必须与横坐标平行地向右移动 k 步，向左也移动 k 步，与纵坐标轴平行地向上移动 l 步，向下也移动 l 步，且 $k+l=n$. 因此

$$u_n=\frac{1}{4^{2n}}\sum_{k=0}^{n}\frac{(2n)!}{[k!\ (n-k)!]^2}=\frac{1}{4^{2n}}\mathrm{C}_{2n}^{n}\sum_{k=0}^{n}(\mathrm{C}_n^k)^2=\frac{1}{4^{2n}}(\mathrm{C}_{2n}^{n})^2\simeq\frac{1}{\pi n},$$

由于 $\sum 1/n=\infty$，平面上的对称的随机游动也是常返的.

我们再讨论空间中的对称随机游动. 这时质点的位置是空间中的整数格点，每个位置有六个相邻的位置. 质点各以 1/6 的概率转移到六个相邻位置中的每一个. 同样地，空间中的对称随机游动也是周期为 2 的不可约链.

质点经过 $2n$ 步返回原位置的概率 u_n 可类似地计算：

$$\begin{aligned}u_n&=\frac{1}{6^{2n}}\sum_{j,k\geqslant0,j+k\leqslant n}\frac{(2n)!}{[j!\ k!\ (n-j-k)!]^2}\\&=\frac{1}{2^{2n}}\mathrm{C}_{2n}^{n}\sum_{j,k\geqslant0,j+k\leqslant n}\left[\frac{1}{3^n}\frac{n!}{j!\ k!\ (n-j-k)!}\right]^2\\&\leqslant\frac{1}{2^{2n}}\mathrm{C}_{2n}^{n}\max_{j,k\geqslant0,j+k\leqslant n}\left[\frac{1}{3^n}\frac{n!}{j!\ k!\ (n-j-k)!}\right].\end{aligned}$$

这里利用了三项式定理：

$$\sum_{j,k\geqslant0,j+k\leqslant n}\left[\frac{1}{3^n}\frac{n!}{j!\ k!\ (n-j-k)!}\right]=1.$$

三项分布的最大项在 j 与 k 最接近 $n/3$ 时达到. 仍由斯特林公式可知,这最大项与 $1/n$ 同阶,从而 u_n 的阶数不超过 $1/n^{3/2}$. 但是 $\sum_n 1/n^{3/2}<\infty$,因此与直线和平面上的对称随机游动不同,空间中的对称随机游动是非常返的.

更一般地,我们可以考虑 $d\geqslant 3$ 维空间中的对称随机游动. 这仍然是一个周期为 2 的不可约链. 可以预期,这时质点经过 $2n$ 步返回原位置的概率 u_n 的阶数不超过 $1/n^{d/2}$. 但是 $\sum_n 1/n^{d/2}<\infty$,因此 $d\geqslant 3$ 维空间中的对称随机游动是非常返的. 然而这是一个在直观上难以理解的特别现象,很难找到直观上可以接受的理由,解释为什么 $d\geqslant 3$ 维空间与一二维空间中的对称随机游动在常返性上会截然不同. 这个结果通常称为波利亚(Polya)定理. 在概率论与数理统计中,$d\geqslant 3$ 维情形的结果与一二维时的结果有本质不同的现象时有发生,令人注目,甚为有趣. □

定理 3.8 设 $\mathrm{P}(r_i<\infty)=1$. 令

$$Y_n=X_{r_i+n},\qquad n\geqslant 1,$$

则 $Y=\{Y_n,n\geqslant 0\}$ 是从 i 出发的具有与 $X=\{X_n,n\geqslant 0\}$ 相同的转移概率矩阵的马尔可夫链,且与 $(r_i,X_0,\cdots,X_{r_i})$ 相互独立.

证 首先注意到 $Y_0=X_{r_i}=i$ 是常量. 对任意的 $n\geqslant 1,k_0\in E,k_1,\cdots,k_{n-1}\in E\setminus\{i\},m\geqslant 1,j_1,\cdots,j_m\in E$,

$$\begin{aligned}
&\mathrm{P}\Big(\bigcap_{l=0}^{n-1}\{X_i=k_l\}\cap\{r_i=n\}\bigcap_{l=1}^{m}\{Y_l=j_l\}\Big)\\
&=\mathrm{P}\Big(\bigcap_{l=0}^{n-1}\{X_i=k_l\}\cap\{X_n=i\}\bigcap_{l=1}^{m}\{X_{n+1}=j_l\}\Big)\\
&=\mathrm{P}\Big(\bigcap_{l=0}^{n-1}\{X_i=k_l\}\cap\{X_n=i\}\Big)\mathrm{P}\Big(\bigcap_{l=1}^{m}\{X_{n+l}=j_l\}\,\Big|\,X_n=i\Big)\\
&=\Big\{\mathrm{P}\Big(\bigcap_{l=0}^{n-1}\{X_i=k_l\}\cap\{X_n=i\}\Big)\Big\}\mathrm{P}_i\Big(\bigcap_{l=1}^{m}\{X_l=j_l\}\Big)\\
&=\Big\{\mathrm{P}\Big(\bigcap_{l=0}^{n-1}\{X_i=k_l\}\cap\{r_i=n\}\Big)\Big\}p_{ij_1}\cdots p_{j_{m-1}j_m}.
\end{aligned}\tag{12}$$

特别,将(12)式对全部的 $n\geqslant 1,i_0\in E,k_1,\cdots,k_{n-1}\in E\setminus\{i\}$ 相加可得

$$\mathrm{P}(Y_1=j_1,\cdots,Y_m=j_m)=p_{ij_1}\cdots p_{j_{m-1}j_m}.\tag{13}$$

(13)即表明 Y 是从 i 出发的马尔可夫链,与 X 有相同的转移概率矩阵,而(12)表明 Y 与 $(r_i,X_0,\cdots,X_{r_i})$ 相互独立. □

定理 3.8 说明马尔可夫链 X 从时刻 r_i 开始,好像又从头开始一样,只是固定从状态 i 出发. 这个结论实质上是 r_i 为 X 的停时这一事实的推论,而马尔可夫链从 r_i 开始与 r_i 之前的状况独立的结论,还得自于 $X_{r_i}=i$ 是常量这一因素.

定理 3.9 设 i 为常返状态. 令

$$r_i(0)=0,\quad r_i(k)=\inf\{n>r_i(k-1):X_n=i\},\quad k\geqslant 1,\tag{14}$$

$$w_i(k)=r_i(k)-r_i(k-1),\qquad k\geqslant 1,\tag{15}$$

则$\{w_i(k),k\geqslant 1\}$为独立随机变量序列，$\{w_i(k),k\geqslant 2\}$同分布，其公共分布列为$\{f_{ii}^{(n)},n\geqslant 1\}$.

证 由状态 i 的常返性，对一切 $k\geqslant 1$，$r_i(k)$为有限停时，它是第 k 次返回状态 i 的时刻. 显然，$r_i(1)=r_i$. 令

$$Y^{(k)}=\{Y_n^{(k)},n\geqslant 0\},\qquad Y_n^{(k)}=X_{r_i(k)+n},\quad n\geqslant 0.$$

易见，$w_i(2)$是 $Y^{(1)}$的首次返回 i 的时刻. 由定理 3.8，$Y^{(1)}$与 X 同转移概率矩阵，与 $w_i(1)=r_i(1)$独立，因此 $w_i(1)$与 $w_i(2)$独立. 实际上，定理 3.8 对每一个 $r_i(k)$适用. $w_i(k+1)$是 $Y^{(k)}$的首次返回 i 的时刻. $Y^{(k)}$与 X 同转移概率分布，与$(X_1,\cdots,X_{r_i(k)})$独立，而$(w_i(1),\cdots,w_i(k))$由$(X_1,\cdots,X_{r_i(k)})$的取值所决定，因此 $w_i(k+1)$与$(w_i(1),\cdots,w_i(k))$独立. 由此即知，$\{w_i(k),k\geqslant 1\}$为独立随机变量序列. 注意到 $Y^{(k)}$都是从 i 出发的，因此$\{w_i(k),k\geqslant 2\}$同分布，我们早已知道 $w_i(2)$的分布是$\{f_{ii}^{(n)},n\geqslant 1\}$. 如果 $\mathsf{P}(X_0=i)=1$，则$\{w_i(k),k\geqslant 1\}$为 i.i.d.. □

习　题

3-1. 设有两个状态的马尔可夫链的转称概率矩阵为

$$\mathbf{P}=\begin{pmatrix}\alpha & 1-\alpha\\ 1-\beta & \beta\end{pmatrix},$$

求$(f_{ij}^{(n)})$，$n\geqslant 1$.

3-2. 设马尔可夫链的转移概率矩阵为

$$(1)\begin{pmatrix}1 & 0\\ 1/2 & 1/2\end{pmatrix};\qquad (2)\begin{bmatrix}1/2 & 1/4 & 1/4\\ 1/2 & 0 & 1/2\\ 1/2 & 1/2 & 0\end{bmatrix};$$

$$(3)\begin{bmatrix}1 & 0 & 0 & 0 & 0 & \cdots\\ 1/2 & 0 & 1/2 & 0 & 0 & \cdots\\ 1/3 & 0 & 0 & 2/3 & 0 & \cdots\\ 1/4 & 0 & 0 & 0 & 3/4 & \cdots\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\end{bmatrix},$$

求矩阵(f_{ij}).

3-3. 证明：对一切 $i,j,k\in E$，

(1) 从 i 出发经过 j 之后再又到达 k 的概率：

$$\mathsf{P}_i\Big(\bigcup_{n=1}^{\infty}\bigcup_{m=n+1}^{\infty}\{X_n=j,X_m=k\}\Big)=f_{ij}f_{jk};$$

(2) $f_{ik}\geqslant f_{ij}f_{jk}$.

3-4. 若在一个本质类中存在两个状态 i,j 使得 $f_{ij}=f_{ji}=1$，则这类是常返的.

3-5. 证明：若对一个不可约链，存在一个状态 k 及正常数 $\alpha>0$，使得对一切 $j\neq k$，$f_{jk}\geqslant\alpha$，则链是常返的.

3-6. 设 j 为常返状态. 证明:对一切 $i\in E$,

$$f_{ij}=\mathrm{P}_i\left(\bigcup_{n=1}^{\infty}\{X_n\in C(j)\}\right).$$

3-7. 设状态 i 的周期为 d. 证明:d 是 $\{n\geqslant 1: f_{ii}^{(n)}>0\}$ 的最大公因子.

3-8. 设马尔可夫链的状态空间为 $\mathbb{Z}$,转移概率为:

$$p_{i,i+a}=p,\qquad p_{i,i-b}=q,\qquad i\in\mathbb{Z},$$
$$p>0,q>0,\ p+q=1,$$

a 与 b 为两个互质的正整数,则马尔可夫链为不可约的,且 $pa=qb$ 时,链为常返的,$pa\neq qb$ 时,链为非常返的.

3-9. 设有限马尔可夫链有 a 个状态,k 为常返状态. 证明:存在常数 q,$0<q<1$,使得 $n>a$ 时,从 k 出发首次返回 k 的时间大于 n 的概率

$$\mathrm{P}_k(\tau_k>n)<q^n.$$

3-10. 设 j 为非常返状态,N_j 为 $\{X_n,n\geqslant 1\}$ 中状态 j 出现的次数. 证明:

$$\mathrm{P}_i(N_j=r)=\begin{cases}f_{ij}(1-f_{jj})f_{jj}^{r-1}, & r\geqslant 1,\\ 1-f_{ij}, & r=0.\end{cases}$$

§2.4 吸收概率与平均吸收时间

在前一节中,我们虽然给出了状态为常返的一个判别法,但这个判别法是利用高阶转移概率的性质来表述的. 然而高阶转移概率的计算本身就是一个困难的课题,所以这个判别法在实际上并不是那么可行的. 同样地,虽然我们有比极限定理可用来计算 f_{ij},但必须先计算高阶转移概率,再求部分和,这也处于同样的困境之中. 在这一节中我们要另辟途径,用解线性方程组的方法来解决这些问题. 对于马尔可夫链,许多概率和数字特征往往是一个非负系数的线性方程组的最小非负解. 因此,非负系数的线性方程组也就成为研究马尔可夫链的一个重要工具. 我们将按照需要,逐步深入讨论非负系数线性方程组的性质.

引理 4.1 设给出线性方程组

$$z_i=\sum_k a_{ik}z_k+b_i,\qquad i\in E,\tag{1}$$

其中 $a_{ik}\geqslant 0,b_i\geqslant 0,i,k\in E$. 定义

$$u_i^{(1)}=b_i,\quad u_i^{(n+1)}=\sum_k a_{ik}u_k^{(n)},\quad n\geqslant 1,\quad i\in E,$$
$$z_i^*=\sum_{n=1}^{\infty}u_i^{(n)},\qquad i\in E,$$

则 $\{z_i^*,i\in E\}$ 是方程组(1)的最小非负解.

证 注意到 $u_i^{(n)}\geqslant 0$,

$$z_i^* = \sum_{n=1}^{\infty} u_i^{(n)} = u_i^{(1)} + \sum_{n=1}^{\infty} u_i^{(n+1)} = b_i + \sum_{n=1}^{\infty} \sum_k a_{ik} u_k^{(n)}$$

$$= b_i + \sum_k a_{ik} \sum_{n=1}^{\infty} u_k^{(n)} = b_i + \sum_k a_{ik} z_k^* .$$

所以$\{z_i^*, i\in E\}$是方程组(1)的非负解.

设$\{\bar{z}_i, i\in E\}$也是方程组(1)的非负解,我们要证明 $\bar{z}_i \geqslant z_i^*, i\in E$. 为此只要证明,对一切 $n\geqslant 1$

$$\bar{z}_i \geqslant z_i^{(n)} = \sum_{v=1}^{n} u_i^{(v)}, \quad i\in E. \tag{2}$$

我们用归纳法证明(2)式. $n=1$ 时是容易看出的:

$$\bar{z}_i = \sum_k a_{ik}\bar{z}_k + b_i \geqslant b_i = u_i^{(1)} = z_i^{(1)}, \qquad i\in E.$$

设(2)式对 n 成立,则

$$\bar{z}_i = \sum_k a_{ik}\bar{z}_k + b_i \geqslant \sum_k a_{ik} z_k^{(n)} + b_i = \sum_k a_{ik} \sum_{v=1}^{n} u_k^{(v)} + u_i^{(1)}$$

$$= \sum_{v=1}^{n} \sum_k a_{ik} u_k^{(v)} + u_i^{(1)} = \sum_{v=1}^{n} u_i^{(v+1)} + u_i^{(1)} = z_i^{(n+1)},$$

因此(2)式对 $n+1$ 也成立. □

在引理 4.1 中,我们没有注意级数的收敛性问题,这是因为这些级数全是正项级数,即使发散到 $+\infty$,也不影响结果的正确性.

将线性方程组(1)写成矩阵的形式更为方便. 记

$$\boldsymbol{Z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \end{pmatrix}, \quad \boldsymbol{A} = (a_{ij}), \quad \boldsymbol{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \end{pmatrix},$$

则方程组(1)可写成

$$\boldsymbol{Z} = \boldsymbol{AZ} + \boldsymbol{B}. \tag{3}$$

现在可用明显表达式写出最小非负解(一向量大于另一向量是指,该向量的分量均大于另一向量的相应的分量). 记

$$\boldsymbol{Z}^* = \begin{pmatrix} z_1^* \\ z_2^* \\ \vdots \end{pmatrix}, \qquad \boldsymbol{U}^{(n)} = \begin{pmatrix} u_1^{(n)} \\ u_2^{(n)} \\ \vdots \end{pmatrix}, \quad n\geqslant 1,$$

则我们有

$$\boldsymbol{U}^{(1)} = \boldsymbol{B}, \quad \boldsymbol{U}^{(n+1)} = \boldsymbol{AU}^{(n)} = \cdots = \boldsymbol{A}^n \boldsymbol{U}^{(1)} = \boldsymbol{A}^n \boldsymbol{B}, \quad n\geqslant 1,$$

$$\boldsymbol{Z}^* = \sum_{n=0}^{\infty} \boldsymbol{A}^n \boldsymbol{B} = \Big(\sum_{n=0}^{\infty} \boldsymbol{A}^n \Big) \boldsymbol{B}. \tag{4}$$

把方程组写成(3)的形式还有一个好处,那就是我们也可以讨论 Z 为矩阵的情形:$Z=(z_{ij})$,这时常数项 B 也是矩阵:$B=(b_{ij})$.换句话说,我们讨论一系列的线性方程组:固定一个 j,就有一个方程组

$$z_{ij}=\sum_{k}a_{ik}z_{kj}+b_{ij},\qquad i\in E.$$

这些方程组仅仅常数项不同.显然,这时最小非负解仍然由(4)式给出(同样地,一个矩阵大于另一个矩阵是指,该矩阵的每个元素均大于另一矩阵的相应元素).

定理 4.1 对任意的 $j\in E$,$\{f_{ij},i\in E\}$是线性方程组

$$z_i=\sum_{k\neq j}p_{ik}z_k+p_{ij},\qquad i\in E \tag{5}$$

的最小非负解.

证 由于 $f_{ij}=\sum_{n=1}^{\infty}f_{ij}^{(n)}$,$f_{ij}^{(1)}=p_{ij}$,如果我们能够证明

$$f_{ij}^{(n+1)}=\sum_{k\neq j}p_{ik}f_{kj}^{(n)},\quad n\geqslant 1, \tag{6}$$

则由引理 4.1 即得定理的结论了.下证(6)式成立.

$$\begin{aligned}
f_{ij}^{(n+1)}&=\mathsf{P}_i(X_{n+1}=j,X_v\neq j,1\leqslant v\leqslant n)\\
&=\sum_{k\neq j}\mathsf{P}_i(X_1=k,X_{n+1}=j,X_v\neq j,2\leqslant v\leqslant n)\\
&=\sum_{k\neq j}\mathsf{P}_i(X_1=k)\mathsf{P}(X_{n+1}=j,X_v\neq j,2\leqslant v\leqslant n\mid X_1=k)\\
&=\sum_{k\neq j}\mathsf{P}_{ik}\mathsf{P}_k(X_n=j,X_v\neq j,1\leqslant v\leqslant n-1)=\sum_{k\neq j}p_{ik}f_{kj}^{(n)}.\quad\square
\end{aligned}$$

推论 不可约链为常返的充要条件是存在 $i\in E$,使对一切 $k\neq i$ 有 $f_{ki}=1$.

证 由定理3.3,若 i,j 属于同一个常返类,则 $f_{ij}=f_{ji}=1$,由此即得必要性.下证充分性.这时由(5)式,

$$f_{ii}=\sum_{k\neq i}p_{ik}f_{ki}+p_{ii}=\sum_{k\neq i}p_{ik}+p_{ii}=\sum_{k}p_{ik}=1.$$

因此链是常返的. □

现在我们对 f_{ij} 作更细致的讨论.设 i 为常返状态,则 $j\in C(i)$时 $f_{ij}=1$,$j\notin C(i)$时 $f_{ij}=0$.所以,只要讨论 i 为非常返状态的情形.

今后以 T 记非常返状态全体.

设 j 为常返状态.对常返的 k,$k\in C(j)$时 $f_{kj}=1$,而 $k\notin C(j)$时 $f_{kj}=0$,因此

$$f_{ij}=\sum_{k\neq j}p_{ik}f_{kj}+p_{ij}=\sum_{k\in T}p_{ik}f_{kj}+\sum_{k\in C(j)}p_{ik},\quad i\in T. \tag{7}$$

设 j 为非常返状态,对常返的 k,$f_{kj}=0$,因此

$$f_{ij}=\sum_{k\neq j}p_{ik}f_{kj}+p_{ij}=\sum_{j\neq k\in T}p_{ik}f_{kj}+p_{ij},\quad i\in T. \tag{8}$$

引理 4.2 设$\{z_i^*,i\in E\}$为线性方程组(1)的最小非负解,$G\subset E$,则线性方程组

$$w_i = \sum_{k \in G} a_{ik} w_k + \left(\sum_{k \notin G} a_{ik} z_k^* + b_i \right), \quad i \in G \tag{9}$$

的最小非负解 $w_i^* = z_i^*, i \in G$.

证 $\{z_i^*, i \in G\}$ 显然是方程组(9)的非负解,因此有 $w_i^* \leqslant z_i^*, i \in G$. 下证相反的不等式成立. 由引理 4.1 的证明可知,只要证明对一切 n

$$w_i^{(n)} \geqslant z_i^{(n)}, \qquad i \in G, \tag{10}$$

这里

$$w_i^{(1)} = b_i, \quad w_i^{(n+1)} = \sum_{k \in G} a_{ik} w_k^{(n)} + \left(\sum_{k \notin G} a_{ik} z_k^* + b_i \right).$$

我们用归纳法证(10)式. $n = 1$ 时

$$w_i^{(1)} = \sum_{k \notin G} a_{ik} z_k^* + b_i \geqslant b_i = z_i^{(1)}$$

是明显的. 设(10)式对 n 成立. 注意到 $z_i^* \geqslant z_i^{(n)}$,我们有

$$\begin{aligned} w_i^{(n+1)} &= \sum_{k \in G} a_{ik} w_k^{(n)} + \sum_{k \notin G} a_{ik} z_k^* + b_i \\ &\geqslant \sum_{k \in G} a_{ik} z_k^{(n)} + \sum_{k \notin G} a_{ik} z_k^{(n)} + b_i = z_i^{(n+1)}. \quad \square \end{aligned}$$

定理 4.2 (1) 设 j 为常返状态,则 $\{f_{ij}, i \in T\}$ 是线性方程组

$$z_i = \sum_{k \in T} p_{ik} z_k + \sum_{k \in C(j)} p_{ik}, \quad i \in T \tag{11}$$

的最小非负解;

(2) 设 j 为非常返状态,则 $\{f_{ij}, i \in T\}$ 是线性方程组

$$z_i = \sum_{j \neq k \in T} p_{ik} z_k + p_{ij}, \quad i \in T \tag{12}$$

的最小非负解.

证 由(7)及(8)式可知,方程组(11)及(12)是在方程组(5)中代入部分最小非负解 $z_k^* = f_{kj}, k \notin T$ 所得,因此由引理 4.2(其中的 G 取为 T)即得定理的结论. $\square$

推论 设状态 j, k 属于同一个常返类,则

$$f_{ij} = f_{ik}, \quad i \in T.$$

证 这时 $\{f_{ij}, i \in T\}$ 及 $\{f_{ik}, i \in T\}$ 是同一个线性方程组(11)的最小非负解,因此相等. $\square$

例 4.1 设转移概率矩阵为

$$\mathbf{P} = \begin{pmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

我们来计算矩阵(f_{ij}). 首先讨论状态的分类：$T=\{0,1\}$，而$\{2,3\}$及$\{4\}$是两个常返类. 由此即知 $f_{20}=f_{21}=0, f_{22}=f_{23}=1, f_{24}=0; f_{30}=f_{31}=0, f_{32}=f_{33}=1, f_{34}=0; f_{40}=f_{41}=f_{42}=f_{43}=0$, $f_{44}=1$. 由方程组(11)

$$f_{12}=\frac{1}{4}f_{12}+\left(\frac{1}{4}+\frac{1}{4}\right),\qquad f_{14}=\frac{1}{4}f_{14}+\frac{1}{4}.$$

因此 $f_{12}=2/3, f_{14}=1/3$. 由定理 4.2 的推论，$f_{13}=f_{12}=2/3$. 再由(11)式

$$f_{02}=\frac{1}{5}f_{02}+\frac{1}{5}f_{12}+\left(\frac{1}{5}+\frac{1}{5}\right),\qquad f_{04}=\frac{1}{5}f_{04}+\frac{1}{5}f_{14}+\frac{1}{5}.$$

因此 $f_{02}=2/3, f_{04}=1/3$. 同理，$f_{03}=f_{02}=2/3$. 注意到从 1 出发或停留原状态，或进入常返状态，即 $1\not\to 0, f_{10}=0$. 由(12)，$f_{11}=1/4$,

$$f_{00}=\frac{1}{5}f_{10}+\frac{1}{5}=\frac{1}{5},\quad f_{01}=\frac{1}{5}f_{01}+\frac{1}{5},\quad f_{01}=\frac{1}{4}.$$

所以

$$(f_{ij})=\begin{pmatrix}1/5 & 1/4 & 2/3 & 2/3 & 1/3\\ 0 & 1/4 & 2/3 & 2/3 & 1/3\\ 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}.\qquad\square$$

例 4.2 有一个吸收壁的随机游动 随机游动的状态空间限制为 $\mathscr{S}=\{0,1,2,\cdots\}$，转移概率为

$$p_{00}=1, p_{i,i+1}=p, p_{i,i-1}=q=1-p, i\geqslant 1, 0<p<1.$$

质点在状态 $i\neq 0$ 时，游动状况与(p,q)随机游动一样. 但状态 0 是吸收状态. 它好比一个吸收壁，质点一旦被它吸进去，游动就结束了.

$j\neq 0$ 时 $j\to 0$，但 $0\not\to j$，因此 $j\neq 0$ 是非本质状态. 若 $i\neq 0, j\neq 0$，则 $i\leftrightarrow j$，因此 $T=\{1,2,\cdots\}$ 是一个非本质类. 我们来求从 $j\neq 0$ 出发最终被状态 0 吸收的吸收概率 f_{j0}. 由定理 4.2，$\{f_{j0}, j\geqslant 1\}$是线性方程组

$$\begin{cases}z_1=pz_2+q,\\ z_j=pz_{j+1}+qz_{j-1}, & j\geqslant 2\end{cases}\tag{13}$$

的最小非负解. 将(13)改写成

$$\begin{cases}p(z_2-z_1)=q(z_1-1),\\ p(z_{j+1}-z_j)=q(z_j-z_{j-1}), & j\geqslant 2,\end{cases}$$

从而可得

$$z_{j+1}-z_j=\frac{q}{p}(z_j-z_{j-1})=\cdots=\left(\frac{q}{p}\right)^j(z_1-1),\quad j\geqslant 2,$$

$$z_{j+1}-1=\sum_{v=0}^{j}\left(\frac{q}{p}\right)^{v}(z_1-1),\quad j\geqslant 0.$$

若 $p\leqslant q$,即 $p\leqslant 1/2$,则 $\sum_{v=0}^{j}(q/p)^{v}\to\infty$,但作为最小非负解的 $\{f_{j0},j\geqslant 1\}$ 是有界的,因此只能 $f_{10}=1$,从而对一切 $j\geqslant 1$,$f_{j0}=1$.若 $p>q$,即 $p>1/2$,由于我们只讨论非负解,

$$0\leqslant\lim_{j\to\infty}z_{j+1}=\lim_{j\to\infty}\left[1+\sum_{v=0}^{j}\left(\frac{q}{p}\right)^{v}(z_1-1)\right]=1+\frac{z_1-1}{1-\frac{q}{p}}.$$

由此可得

$$z_1-1\geqslant-\left(1-\frac{q}{p}\right),\quad z_1\geqslant\frac{q}{p},$$

$$z_{j+1}\geqslant 1-\sum_{v=0}^{j}\left(\frac{q}{p}\right)^{v}\left(1-\frac{q}{p}\right)=\left(\frac{q}{p}\right)^{j+1},\quad j\geqslant 0.$$

另一方面,不难直接验证 $\{z_j=(q/p)^j,j\geqslant 1\}$ 是方程组(13)的解,所以它就是最小非负解,即 $f_{j0}=(q/p)^j,j\geqslant 1$.所得结果的直观意义也很明确.当 $p\leqslant 1/2$ 时,从任何一个状态 $j\geqslant 1$ 出发,迟早要被状态 0 吸收:

$$\mathrm{P}_j(\lim_{n\to\infty}X_n=0)=1.$$

这里 $p=1/2$ 时的结论单凭直观的猜测是难以想到的.当 $p>1/2$ 时,如果质点的初始位置 $j\geqslant 1$,则

$$\mathrm{P}_j(\lim_{n\to\infty}X_n=0)=\left(\frac{q}{p}\right)^{j}.$$

另一方面,T 是非常返的:从 j 出发到达任何一个状态 $k\geqslant 1$ 至多有限次.因此如果质点不被状态 0 吸收,对任意 $k\geqslant 1$,质点迟早要永远离开 $\{1,2,\cdots,k\}$,也就是要向无穷远处跑去:

$$\mathrm{P}_j(\lim_{n\to\infty}X_n=\infty)=1-\left(\frac{q}{p}\right)^{j}.\qquad\square$$

例 4.3 有两个吸收壁的随机游动 随机游动的状态空间限制为 $\{0,1,\cdots,a-1,a\}$,转移概率为

$$p_{00}=p_{aa}=1,$$

$$p_{i,i+1}=p,p_{i,i-1}=q=1-p,1\leqslant i\leqslant a-1,0<p<1.$$

现在 0 与 1 是两个吸收状态,$T=\{1,2,\cdots,a-1\}$ 为非本质类.我们仍来求吸收概率 $f_{j0},f_{ja},1\leqslant j\leqslant a-1$.由有限链的性质(定理 3.6 的推论 2)可知 $f_{j0}+f_{ja}=1$,因此只需求 f_{j0} 即可.由定理 4.2,只要解线性方程组

$$\begin{cases}z_1=pz_2+q,\\ z_j=pz_{j+1}+qz_{j-1},\qquad 2\leqslant j\leqslant a-2,\\ z_{a-1}=qz_{a-2},\end{cases}\tag{14}$$

如果定义 $z_0=1(=f_{00}), z_a=0(=f_{a0})$，则方程组(14)可统一地写成

$$z_j=pz_{j+1}+qz_{j-1}, \qquad 1\leqslant j\leqslant a-1.$$

解这个方程组并不困难：

$$p(z_{j+1}-z_j)=q(z_j-z_{j-1}),$$

$$z_{j+1}-z_j=\left(\frac{q}{p}\right)(z_j-z_{j-1})=\cdots=\left(\frac{q}{p}\right)^j(z_1-z_0),$$

$$z_{j+1}-1=\sum_{v=0}^{j}\left(\frac{q}{p}\right)^v(z_1-1), 1\leqslant j\leqslant a-1.$$

由 $z_0=0$ 解得

$$1-z_1=\left\{\sum_{v=0}^{a-1}\left(\frac{q}{p}\right)^v\right\}^{-1},$$

$$f_{j0}=z_j=\begin{cases}1-\dfrac{j}{a}, & p=\dfrac{1}{2},\\ \dfrac{\left(\dfrac{q}{p}\right)^a-\left(\dfrac{q}{p}\right)^j}{\left(\dfrac{q}{p}\right)^a-1}, & p\neq\dfrac{1}{2},\end{cases} \qquad 1\leqslant j\leqslant a-1.$$

容易看出，当 $a\to\infty$ 时，上式中的 f_{j0} 的极限就是例 4.2 中所求得的结果. 这是合乎情理的，因为把右面的吸收壁移动到无穷远处，例 4.3 就化成例 4.2 了.

这个例子所讨论的问题在很多概率论教科书中都可找到. 一方面，这是因为方程组(14)可以直接用全概率公式简单地推导出来，理解这问题也不需要了解马尔可夫链的概念. 另一方面，还因为这个问题就是概率论历史上一个极为著名的问题——“赌徒输光问题”. 它原来的叙述是这样的：甲、乙两个赌徒进行一系列的赌博. 在每一局中甲获胜的概率为 p，乙获胜的概率为 q. 假设各局赌博的结果是相互独立的. 每一局后，失败者付给获胜者一元钱. 假设开始时甲有 b 元，乙有 c 元. 赌博一直进行到有一个人全部输光为止. 求甲或乙输光的概率各是多少？记 $a=b+c$. 以 X_n 表示赌了 n 局之后甲手中的赌金，那么 $\{X_n, n\geqslant 0\}$ 就是我们上面讨论的有两个吸收壁的随机游动，且游动从 $X_0=b$ 出发. 质点被状态 0 吸收表示甲输光，被状态 a 吸收则表示乙输光. 因此甲及乙输光的概率即分别为 f_{b0} 及 f_{ba}.

分析一下所得的结果是十分有意思的. 当 $p=q=1/2$ 时，即甲乙两人的赌博本领相当时，甲、乙两人输光的概率分别为 $\frac{c}{b+c}$ 及 $\frac{b}{b+c}$. 显然，谁的初始赌本大，谁就处于有利的地位. 但 $p\neq q$ 时，一下子还无法看清究竟是 p 大还是 b 大起主要作用. 为此我们给出一些数值的结果. 下表是从其中一个赌徒的观点来计算的.

开始的赌本	对手的赌本	每局获胜的概率	输光的概率
90	10	0.50	0.100
90	10	0.45	0.866
90	10	0.40	0.983
99	1	0.40	0.333

从上表中可看出,尽管在赌本上占有明显的优势,但本领差的话,还会落到一个糟糕的结局.这无疑给了我们一个启示:有高超的技术比拥有雄厚的资本更为重要. □

例 4.4 分支过程 考虑一个群体的"繁殖"问题.开始时群体所含个体数记为 X_0,它们称为第零代的个体.第零代个体的后代为第一代,第一代的个体个数记为 X_1.这样继续下去,一般地,第 n 代是第 $n-1$ 代的后代,第 n 代个体的个数记为 X_n.我们假设,同一代中各个个体所产生的后代个数是相互独立的,且与群体以前的繁殖过程无关.每一个体所产生的后代个数 ξ 是随机变量,且分布为

$$\mathrm{P}(\xi=k)=g_k, \quad k\geqslant 0. \tag{15}$$

下面就是这个繁殖模型的严格的数学描述.

设 $\{\xi_{nm}, n, m=1,2,\cdots\}$ 为独立同分布的取非负整数值的随机变量,其共同分布由(15)给出.X_0 是一个与 $\{\xi_{nm}, n, m=1,2,\cdots\}$ 独立的取非负整数值的随机变量,$\{X_n, n\geqslant 1\}$ 按下式递归地定义:

$$X_n=\begin{cases}\xi_{n1}+\cdots+\xi_{n,X_{n-1}}, & \text{若 } X_{n-1}\geqslant 1,\\ 0, & \text{若 } X_{n-1}=0,\end{cases} \quad n\geqslant 1. \tag{16}$$

由此可知,若已知 $X_{n-1}=i\geqslant 1$,则 X_n 的分布是 i 个独立同分布随机变量之和的分布(容易看出,$\{X_0,\cdots,X_{n-1}\}$ 与 $\{\xi_{n1},\xi_{n2},\cdots\}$ 相互独立),且共同分布由(15)给出.若已知 $X_{n-1}=0$,则 $X_n=0$.因此 X_n 的分布只与 X_{n-1} 的值有关,与 $X_0, X_1,\cdots,X_{n-2}$ 无关.所以 $\{X_n, n\geqslant 0\}$ 为马尔可夫链.这种形式的马尔可夫链就称为分支过程.

为了决定分支过程的转移概率,我们要利用概率分布的母函数这一数学工具.

引理 4.3 设 $\eta,\zeta_1,\zeta_2,\cdots$ 为相互独立的只取非负整数值的随机变量,$\{\zeta_n, n\geqslant 1\}$ 具有共同的分布,其母函数记为 $H(z)$.定义

$$\zeta=\begin{cases}\zeta_1+\zeta_2+\cdots+\zeta_\eta, & \text{若 } \eta\geqslant 1,\\ 0, & \text{若 } \eta=0,\end{cases}$$

则 ζ 的母函数为

$$G_\zeta(z)=G_\eta(H(z)).$$

证 当 $v=0$ 时,理解 $\zeta_1+\cdots+\zeta_v$ 为零,那么

$$\begin{aligned} G_\zeta(z) &= \sum_{k=0}^{\infty}\mathrm{P}(\zeta=k)z^k \\ &= \sum_{k=0}^{\infty}\sum_{v=0}^{\infty}\mathrm{P}(\eta=v,\zeta_1+\cdots+\zeta_v=k)z^k \\ &= \sum_{v=0}^{\infty}\mathrm{P}(\eta=v)\sum_{k=0}^{\infty}\mathrm{P}(\zeta_1+\cdots+\zeta_v=k)z^k. \end{aligned}$$

但

$$\sum_{k=0}^{\infty}\mathrm{P}(\zeta_1+\cdots+\zeta_v=k)z^k=\mathrm{E}\left[z^{\zeta_1+\cdots+\zeta_v}\right]$$

$$= \prod_{j=1}^{v} \mathrm{E}[z^{\zeta_j}] = [H(z)]^v, \quad v \geqslant 0.$$

从而得到

$$G_\zeta(z) = \sum_{v=0}^{\infty} \mathrm{P}(\eta = v)[H(z)]^v = G_\eta(H(z)). \quad \square$$

利用上述引理,我们就能写出分支过程的高阶转移概率的母函数了.记

$$G(z) = \sum_{k=0}^{\infty} g_k z^k.$$

显然,$G(z)$是确定分支过程性质的最基本的数字特征.记

$$F_{in}(z) = \sum_{k=0}^{\infty} p_{ik}^{(n)} z^k.$$

设 $X_0 = 1$,则 $F_{1n}(z)$即为 X_n 的母函数.现在由定义式(16)及引理 4.3 可知

$$\begin{aligned} F_{1n}(z) &= F_{1,n-1}(G(z)) = F_{1,n-2}(G(G(z))) = \cdots \\ &= F_{10}(G^{(n)}(z)) = G^{(n)}(z), \end{aligned} \tag{17}$$

这里 $F_{10}(z) = \sum_{k=0}^{\infty} p_{1k}^{(0)} z^k = z, G^{(n)}(z) = \underbrace{G(G(\cdots G}_{n\text{次}}(z))\cdots)$.

设 $X_0 = i \geqslant 1$,则 $F_{in}(z)$为 X_n 的母函数.这时 X_n 是i 个第零代的个体所衍生的第 n 代个体之和.按照假定,这 i 个第零代的个体繁殖后代的情况是相互独立的,而每一个第零代个体所衍生第 n 代个体数的母函数为$F_{1n}(z)$,因此我们有

$$F_{in}(z) = [F_{1n}(z)]^i = [G^{(n)}(z)]^i. \tag{18}$$

这样,由 $G(z)$及(17),(18)式完全确定了分支过程的转移概率,虽然要给出 $p_{ik}^{(n)}$的明显表达式是困难的.

作为(18)式的一个应用,若 $m = \sum_{n=1}^{\infty} k g_k = G'(1) < \infty$,则

$$[G^{(n)}(z)]'|_{z=1} = G'(1)[G^{(n-1)}(z)]'|_{z=1} = \cdots = [G'(1)]^n = m^n,$$

$$\mathrm{E}[X_n | X_0 = i] = F_{in}'(1) = i m^n.$$

$\{X_n = 0\}$表示群体在第 n 代已灭绝.显然,“0”是一个吸收状态.今后我们总假设 $0 < g_0 < 1$.不然的话,$g_0 = 0$ 意味着每个个体至少有一个后代,只要 $g_1 < 1$,群体总数只增不减,必定趋于无穷(每个个体始终只生一个后代的概率随 g_1^n 趋于零而为零).但 $g_1 = 1$ 的情形是毋需讨论的平凡情形,这时群体的总数在每一代都保持为定数.$g_0 = 1$ 是毫无意义的,因为这时 $X_n = 0, n \geqslant 1$.现在由(18)可知 $i \neq 0$ 时

$$p_{i0} = F_{i1}(0) = [F_{11}(0)]^i = [G(0)]^i = g_0^i > 0,$$

即 $i \to 0$,但 $0 \not\to i$,所以 $i \neq 0$ 都是非本质状态,“0”是唯一的吸收状态.对于这样的马尔可夫链,首要的自然是讨论从状态 $i \neq 0$ 出发被状态“0”吸收的概率 f_{i0},也就是开始时群体有 i 个个体,但

迟早要灭绝的概率.这就是分支过程的灭种问题.

由于状态"0"是吸收状态,$f_{i0}=\lim_{n\to\infty}p_{i0}^{(n)}$($\{p_{i0}^{(n)}\}$单调上升).由(18),$p_{i0}^{(n)}=F_{in}(0)=[F_{1n}(0)]^i=(p_{10}^{(n)})^i$,故得

$$f_{i0}=(f_{10})^i,\qquad i\geqslant 1.$$

所以只要讨论 f_{10}就行了,f_{10}就称为分支过程的灭种概率.

引理4.4 设 $F(z)$为概率分布$\{f_k,k\geqslant 0\}$的母函数:$F(z)=\sum_{k=0}^{\infty}f_kz^k$,$0<f_0<1$,则方程

$$z=F(z)\tag{19}$$

有最小正根 z^*.此外,

1) $m=\sum_{k=1}^{\infty}kf_k\leqslant 1$ 时,$z^*=1$;

2) $m>1$ 时,$z^*<1$,且 z^*是这方程在$(0,1)$中的唯一的根.

证 令 $z_1=F(0)=f_0$,$z_{n+1}=F(z_n)$,$n\geqslant 1$,则

$$0<z_n\leqslant z_{n+1}\leqslant 1,\quad n\geqslant 1.\tag{20}$$

不难用归纳法证(20)式.$0<z_1=f_0<1$,$z_1=F(0)\leqslant F(z_1)=z_2\leqslant F(1)=1$($F(z)$在$[0,1]$上为单调上升连续函数),因此 $n=1$ 时(20)式成立.设(20)式对 n 成立,则

$$z_{n+1}=F(z_n)\leqslant F(z_{n+1})=z_{n+2}\leqslant F(1)=1,$$

即(20)式对 $n+1$ 也成立.

记 $z^*=\lim_{n\to\infty}z_n$,则 $z^*=F(z^*)$,$z^*\geqslant f_0>0$,所以 z^*为方程(19)的正根.设 $\bar{z}$ 也为这方程的正根,则

$$\bar{z}=F(\bar{z})\geqslant F(0)=z_1.$$

用归纳法易证 $\bar{z}\geqslant z_n$.事实上,若 $\bar{z}\geqslant z_n$,则 $\bar{z}=F(\bar{z})\geqslant F(z_n)=z_{n+1}$.由此即得 $\bar{z}\geqslant z^*$,所以 z^*为方程(19)的最小正根.

先讨论 $f_0+f_1=1$ 的情形.这时 $m=f_1<1$,且由直接计算可知(19)只有唯一的根 $z^*=1$.

现在讨论 $0<f_0+f_1<1$ 的情形.这时 $F''(z)=2f_2+6f_3+\cdots$在$(0,1)$上恒为正,因此 $F'(z)$在$[0,1]$上为严格单调上升函数.此时又分两种情况如下:

当 $m\leqslant 1$ 时,即 $F'(1)=m\leqslant 1$.由于 $F'(z)$在$[0,1]$上严格单调上升,所以 $z\in(0,1)$时 $F'(z)<1$.由中值定理,对任一 $z\in(0,1)$,存在 $w\in(z,1)$使

$$F(z)-F(1)=F'(w)(z-1)>z-1.$$

由于 $F(1)=1$,即得 $F(z)>z$.所以方程(19)在$(0,1)$中无根,但 1 总是这方程的根,因此 $z^*=1$.

当 $m>1$ 时,由 $F'(z)$的连续性,存在 $z_0\in(0,1)$,使 $z\in(z_0,1)$时 $F'(z)>1$.同样由中值定理,对任一 $z_1\in(z_0,1)$,存在 $w\in(z_1,1)$,使得

$$F(z_1)-F(1)=F'(w)(z_1-1)<z_1-1,$$

即得 $F(z_1)-z_1<0$. 但 $F(0)-0=f_0>0$，因此方程(19)在$(0,z_1)$中有根，从而 $z^*<1$. 若(19)在(0,1)中有两个根：$0<z_1<z_2<1$，则由罗尔(Rolle)定理，有 $w_1\in(z_1,z_2)$，$w_2\in(z_2,1)$，使 $F'(w_1)=F'(w_2)=1$，再由罗尔定理又有 $v\in(w_1,w_2)$，使 $F''(v)=0$，与 $F''(z)$在(0,1)上恒为正矛盾. 所以 z^* 是方程(19)在(0,1)中的唯一的根. □

现在将引理 4.4 用于灭种概率的讨论. 注意到 $p_{i0}^{(n)}=(p_{10}^{(n)})^i$，

$$p_{10}^{(1)}=g_0=G(0),$$

$$\begin{aligned} p_{10}^{(n+1)}=\mathsf{P}_1(X_{n+1}=0) &= \sum_{k=0}^{\infty}\mathsf{P}_1(X_1=k,X_{n+1}=0)\\ &= \sum_{k=0}^{\infty}p_{1k}p_{k0}^{(n)}=\sum_{k=0}^{\infty}g_k(p_{10}^{(n)})^k=G(p_{10}^{(n)}),\end{aligned}$$

因此，灭种概率 $f_{10}=\lim_{n\to\infty}p_{10}^{(n)}$正是方程 $z=G(z)$的最小正根. $m=\sum_{k=1}^{\infty}kg_k$ 是一个个体产生的后代个数的平均数. 当 $m\leqslant 1$ 时，$f_{10}=1$，从而对一切 $i\geqslant 1$，$f_{i0}=1$，即群体迟早必定要灭绝(这事实在 $m=1$ 时直观上并不明显). 当 $m>1$ 时，群体以$(f_{10})^i$ 的概率灭绝(i 为开始群体中个体的个数)，即

$$\mathsf{P}_i(\lim_{n\to\infty}X_n=0)=(f_{10})^i.$$

注意到 $i\neq 0$ 都是滑过状态，就像讨论有一个吸收壁的随机游动时一样，我们有

$$\mathsf{P}_i(\lim_{n\to\infty}X_n=\infty)=1-(f_{10})^i.$$

当 f_{10}很小，i 很大时，$1-(f_{10})^i$ 十分接近于 1，群体中个体个数几乎总要变成无穷. 这可以作为原子核连锁反应中导致爆炸的数学说明. 事实上，可以用分支过程近似地刻画原子核连锁反应. 一个中子与裂变物质发生随机碰撞，产生多个中子，产生 k 个中子的概率为 g_k，$k\geqslant 1$. g_0 是一个中子不与裂变物质碰撞而湮灭的概率. 一次裂变产生的中子平均数 m 很大，因而灭种概率 f_{10}很小. 在引爆原子弹时，使初始的中子数 i 很大，从而导致爆炸. □

现在我们回到方程组(11)上来. 在例 4.3 中，这方程组有唯一的解，但在例 4.2 中，这方程组的解不唯一. 下面我们就来讨论这方程组有唯一解的条件. 这等价于讨论齐次方程组

$$\begin{cases} z_i=\sum_k p_{ik}z_k, \\ |z_i|\leqslant 1, \end{cases}\qquad i\in T \tag{21}$$

的解全为零的条件.

引理 4.5 设给出齐次线性方程组

$$\begin{cases} z_i=\sum_k a_{ik}z_k, \\ |z_i|\leqslant 1, \end{cases}\qquad i\in E, \tag{22}$$

其中 $a_{ik}\geqslant 0$，$\sum_k a_{ik}\leqslant 1$，$i,k\in E$. 定义

$$u_i^{(1)}=\sum_k a_{ik},\ u_i^{(n+1)}=\sum_k a_{ik}u_k^{(n)},\ n\geqslant 1,\quad i\in E,$$

则 $u_i^{(n)} \downarrow u_i$，且$\{u_i, i\in E\}$是方程组(22)的解.

设$\{u'_i, i\in E\}$也为方程组(22)的解，则

$$|u'_i| \leqslant u_i, \qquad i\in E$$

(因此$\{u_i, i\in E\}$称为方程组(22)的最大解).

证 先用归纳法证明 $u_i^{(n)}$ 单调下降. 由假设 $u_i^{(1)}\leqslant 1$，因此

$$u_i^{(2)} = \sum_k a_{ik}u_k^{(1)} \leqslant \sum_k a_{ik} = u_i^{(1)}, \quad i\in E.$$

设 $u_i^{(n)} \leqslant u_i^{(n-1)}, i\in E$，则

$$u_i^{(n+1)} = \sum_k a_{ik}u_k^{(n)} \leqslant \sum_k a_{ik}u_k^{(n-1)} = u_i^{(n)}, \quad i\in E.$$

由此即知 $u_i^{(n)} \downarrow u_i, i\in E$. 注意到级数 $\sum_k a_{ik}u_k^{(n)}$ 关于 n 一致收敛，在 $u_i^{(n+1)}$ 的定义式中取极限 $n\to\infty$，即知$\{u_i, i\in E\}$是方程组(22)的非负解.

对方程组(22)的另一组解$\{u'_i, i\in E\}$，我们用归纳法证明，对一切 n，

$$|u'_i| \leqslant u_i^{(n)}, \quad i\in E.$$

因为$|u'_i|\leqslant 1, i\in E$，所以

$$|u'_i| \leqslant \sum_k a_{ik}|u'_k| \leqslant \sum_k a_{ik} = u_i^{(1)}, \quad i\in E.$$

设$|u'_i|\leqslant u_i^{(n)}, i\in E$，则

$$|u'_i| \leqslant \sum_k a_{ik}|u'_k| \leqslant \sum_k a_{ik}u_k^{(n)} = u_i^{(n+1)}, \quad i\in E.$$

由此即得$|u'_i|\leqslant u_i, i\in E$. □

定理 4.3 设 v_i 为从状态 $i\in T$ 出发，但永远停留在 T 中的概率：

$$v_i = \mathsf{P}_i\left(\bigcap_{n=1}^{\infty}\{X_n\in T\}\right), \quad i\in T,$$

则$\{v_i, i\in T\}$为方程组(21)的最大解.

证 记 $v_i^{(n)} = \mathsf{P}_i\left(\bigcap_{\nu=1}^{n}\{X_\nu\in T\}\right), i\in T$. 易见

$$v_i^{(1)} = \sum_{k\in T} p_{ik} \leqslant 1,$$

$$v_i^{(n+1)} = \mathsf{P}_i\left(\bigcap_{\nu=1}^{n+1}\{X_\nu\in T\}\right) = \sum_{k\in T}\mathsf{P}_i(X_1=k, X_2\in T, \cdots, X_{n+1}\in T)$$

$$= \sum_{k\in T} p_{ik}\mathsf{P}\left(\bigcap_{\nu=2}^{n+1}\{X_\nu\in T\} \,\Big|\, X_1=k\right) = \sum_{k\in T} p_{ik}v_k^{(n)}, \quad i\in T.$$

由引理 4.5 即知，$\{v_i, i\in T\}$是方程组(21)的最大解. □

推论 方程组(11)有唯一有界解的充要条件是对一切 $i\in T, v_i=0$. 特别，对有限马尔可夫链，方程组(11)有唯一有界解.

讨论方程组(11)是否有唯一有界解是一个代数问题，而验证$\{v_i, i\in T\}$是否全为零，对于状

态个数无限的马尔可夫链，一般却是一个困难的概率问题. 恰恰为了知道$\{v_i, i\in T\}$是否全为零，我们反而去判断方程组(11)是否有唯一有界解，或方程组(21)是否只有零解. 按照这样的思路，我们可得到一个极为有用的常返判别法.

定理 4.4 不可约链为常返的充要条件是存在一个(或全部的)$j\in E$使线性方程组

$$z_i = \sum_{k\neq j} p_{ik} z_k, \quad i\neq j \tag{23}$$

的有界解全为零.

证 我们考虑一个新的马尔可夫链，在新的链中状态j改成吸收状态，而其它的转移概率$p_{ik}, i\neq j, k\in E$都保持不变. 从状态$i(\neq j)$出发，直至首次到达j，新的链与原来的链并无不同. 因此$f_{ij}, i\neq j$保持不变. 只是在新的链中，它变成从i出发被j吸收的概率. 不难看出，若$i\neq j$，在新的链中，$i\to j, j\not\to i$，故$T=\{i: i\neq j\}, v_i = 1 - f_{ij}, i\in T$. 由定理 4.3 的推论，方程组(23)的有界解全为零等价于$f_{ij}=1, i\neq j$. 由定理 4.1 的推论，这即等价于链为常返的. 从上述证明可看出，对常返不可约链任取$j\in E$，方程组(23)的有界解全为零. □

定理 4.4 的证明提供了一个求f_{ij}的十分有用的技巧，这就是为了求f_{ij}不妨可以认为j是吸收状态，更进一步，如果在两个马尔可夫链中，从状态i出发直至首次到达状态j的情况完全相同，只是在到达j之后情况才有所不同，那么对这两个马尔可夫链，f_{ij}是相同的. 这虽然是一个显而易见的道理，但在许多场合，却能给我们带来很大的方便. 特别，借助于这个方法，我们可以利用已知的一些马尔可夫链中关于f_{ij}的结果来推算另一些马尔可夫链的f_{ij}.

将方程组(23)略作变形，可得下述形式的常返判别法，在有些场合，它使用起来更方便些.

定理 4.5 不可约链为常返的充要条件是存在一个(或全部的)$j\in E$，使线性方程组

$$z_i = \sum_{k} p_{ik} z_k, \quad i\neq j \tag{24}$$

的有界解必为常数.

证 设方程组(24)有非常数的有界解$\{u_i, i\in E\}$. 令

$$w_i = u_i - u_j, \quad i\neq j,$$

则$\{w_i, i\neq j\}$为非零有界数列，且$i\neq j$时

$$\sum_{k\neq j} p_{ik} w_k = \sum_{k\neq j} p_{ik} u_k - (1-p_{ij}) u_j = \sum_{k} p_{ik} u_k - u_j = u_i - u_j = w_i,$$

即$\{w_i, i\neq j\}$为方程组(23)的非零有界解.

反之，设方程组(23)有非零有界解$\{u_i, i\neq j\}$. 令

$$w_i = \begin{cases} u_i, & i\neq j, \\ 0, & i = j, \end{cases}$$

则$\{w_i, i\in E\}$为非常数有界数列，且$i\neq j$时

$$\sum_{k} p_{ik} w_k = \sum_{k\neq j} p_{ik} u_k = u_i = w_i,$$

即$\{w_i, i\in E\}$为方程组(24)的非常数有界解. □

定义 线性不等式组

$$z_i \geqslant \sum_k p_{ik} z_k, \quad i \in E \tag{25}$$

的非负解称为马尔可夫链的过分函数.

引理 4.6 不可约链的过分函数或恒为零,或恒为正.

证 设$\{h_i, i\in E\}$为过分函数,且不全为零,则有 j 使$h_j>0$.用归纳法及科尔莫戈罗夫-查普曼方程易证,过分函数满足不等式组

$$h_i \geqslant \sum_k p_{ik}^{(n)} h_k, \qquad n\geqslant 1,\ i\in E.$$

对任一 $i\in E$, $i\to j$,故有 $n\geqslant 1$,使 $p_{ij}^{(n)}>0$.因此

$$h_i \geqslant p_{ij}^{(n)} h_j > 0. \quad \square$$

定理 4.6 不可约链为常返的充要条件是过分函数必为常数.

证 先证必要性.设$\{h_i, i\in E\}$为过分函数,全为零的情形显然不必讨论.现在取定 j,令 $u_i=\dfrac{h_i}{h_j}$, $i\in E$,则

$$u_i \geqslant \sum_k p_{ik} u_k = \sum_{k\neq j} p_{ik} u_k + p_{ij} = \sum_{k\neq j} p_{ik} u_k + f_{ij}^{(1)}. \tag{26}$$

在(26)中作迭代:

$$\begin{aligned} u_i &\geqslant \sum_{k\neq j} p_{ik}\Big(\sum_{l\neq j} p_{kl} u_l + p_{kj}\Big) + p_{ij} \\ &= \sum_{k\neq j, l\neq j} p_{ik} p_{kl} u_l + \sum_{k\neq j} p_{ik} p_{kj} + p_{ij} \\ &= \sum_{k\neq j, l\neq j} p_{ik} p_{kl} u_l + f_{ij}^{(2)} + f_{ij}^{(1)}, \end{aligned}$$

继续作迭代,由归纳法可知对一切 $n\geqslant 1$,

$$u_i \geqslant f_{ij}^{(n)} + \cdots + f_{ij}^{(2)} + f_{ij}^{(1)},$$

从而 $u_i\geqslant f_{ij}=1$, $h_i\geqslant h_j$.但 i,j 是任意取的,因此必有 $h_i=h_j$, $i,j\in E$,即$\{h_i, i\in E\}$为常数.

再证充分性.任取 $j\in E$,定义

$$h_i = \begin{cases} f_{ij}, & i\neq j, \\ 1, & i=j, \end{cases}$$

则 $i\neq j$ 时

$$h_i = f_{ij} = \sum_{k\neq j} p_{ik} f_{kj} + p_{ij} = \sum_k p_{ik} h_k,$$

$$h_j = 1 \geqslant f_{jj} = \sum_{k\neq j} p_{jk} f_{kj} + p_{jj} = \sum_k p_{jk} h_k.$$

因此$\{h_i, i\in E\}$为过分函数,必须为常数,即 $i\neq j$ 时, $f_{ij}=1$.由定理 4.1 的推论即知链是常返的. $\square$

例 4.5 有一个反射壁的随机游动 随机游动的状态空间为$\mathbb{N}$,转移概率矩阵为

$$\mathbf{P}=\begin{pmatrix} r_0 & p_0 & 0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & 0 & \cdots \\ 0 & q_2 & r_2 & p_2 & 0 & \cdots \\ 0 & 0 & q_3 & r_3 & p_3 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

其中 $p_i>0, i\geqslant 0; q_i>0, i\geqslant 1; r_0+p_0=1, q_i+r_i+p_i=1, i\geqslant 1$. 容易看出,对任一 $i\geqslant 0, i\leftrightarrow i+1$,因此链是不可约的. 在 $r_0>0$ 时,我们可以设想从状态"0"出发,质点仍可能向左移动,但在"$-1/2$"点处有一座反射壁,质点碰上去被反射回原状态"0". 为了考察链是否为常返,讨论定理4.5中的方程组(24)

$$z_i=q_iz_{i-1}+r_iz_i+p_iz_{i+1},\quad i\geqslant 1. \tag{27}$$

注意到 $1-r_i=p_i+q_i$,因此(27)可写成

$$p_i(z_{i+1}-z_i)=q_i(z_i-z_{i-1}),$$

$$z_{i+1}-z_i=\frac{q_i}{p_i}(z_i-z_{i-1})=\cdots=\frac{q_iq_{i-1}\cdots q_1}{p_ip_{i-1}\cdots p_1}(z_1-z_0).$$

由此即得

$$z_{i+1}=z_1+\left(\sum_{k=1}^{i}\frac{q_1\cdots q_k}{p_1\cdots p_k}\right)(z_1-z_0),\quad i\geqslant 1.$$

z_0 及 z_1 的选取是任意的. 为了使 $\{z_i, i\geqslant 0\}$ 为非常数有界数列,必须且只需

$$z_1\neq z_0 \quad 及 \quad \sum_{k=1}^{\infty}\frac{q_1\cdots q_k}{p_1\cdots p_k}<\infty.$$

由此可见,链为常返的充要条件是

$$\sum_{k=1}^{\infty}\frac{q_1\cdots q_k}{p_1\cdots p_k}=\infty.$$

在 $p_i=p, q_i=q, i\geqslant 1$ 时,

1) 若 $p\leqslant q$,链为常返的;

2) 若 $p>q$,链为非常返的. □

设 $f_{ij}=1$,则 $\{f_{ij}^{(n)}, n\geqslant 1\}$ 是从状态 i 出发首次到达状态 j 的时间 τ_j 的概率分布,所以

$$m_{ij}=\sum_{n=1}^{\infty}nf_{ij}^{(n)} \tag{28}$$

是从 i 首次到达 j 的平均时间: $\mathrm{E}_i[\tau_j]$. 若 $f_{ij}<1$,这时按(28)式定义的 m_{ij} 就没有这个概率意义了. 下面我们来讨论如何求得 m_{ij},而不问 f_{ij} 是等于1,还是小于1.

引理 4.7 设给出线性方程组

$$z_i=\sum_k a_{ik}z_k+b_i,\quad i\in E, \tag{29}$$

其中 $a_{ik}\geqslant 0, b_i\geqslant 0, i,k\in E$. 令

$$b_i=\sum_{n=1}^{\infty} b_i^{(n)},\ b_i^{(n)}\geqslant 0,\quad n\geqslant 1,\ i\in E,$$

$$\bar{u}_i^{(1)}=b_i^{(1)},\bar{u}_i^{(n+1)}=\sum_k a_{ik}\bar{u}_k^{(n)}+b_i^{(n+1)},\quad n\geqslant 1,\ i\in E,$$

$$\bar{z}_i=\sum_{n=1}^{\infty}\bar{u}_i^{(n)},\quad i\in E,$$

则 $\{\bar{z}_i, i\in E\}$ 为方程组(29)的最小非负解.

证 $i\in E$ 时,

$$\begin{aligned}\bar{z}_i&=\sum_{n=1}^{\infty}\bar{u}_i^{(n)}=b_i^{(1)}+\sum_{n=1}^{\infty}\bar{u}_i^{(n+1)}=b_i^{(1)}+\sum_{n=1}^{\infty}\left(\sum_k a_{ik}\bar{u}_k^{(n)}+b_i^{(n+1)}\right)\\&=\sum_k a_{ik}\sum_{n=1}^{\infty}\bar{u}_k^{(n)}+\sum_{n=1}^{\infty}b_i^{(n)}=\sum_k a_{ik}\bar{z}_k+b_i.\end{aligned}$$

所以 $\{\bar{z}_i, i\in E\}$ 是方程组(29)的非负解.

设 $\{z_i^*, i\in E\}$ 为方程组(29)的最小非负解. 我们采用引理 4.1 中的记号. 只要用归纳法证明对一切 $n\geqslant 1$,

$$\sum_{\upsilon=1}^{n}\bar{u}_i^{(\upsilon)}\leqslant\sum_{\upsilon=1}^{n}u_i^{(\upsilon)},\quad i\in E,\tag{30}$$

即知 $\bar{z}_i\leqslant z_i^*$, 从而 $\bar{z}_i=z_i^*$, $i\in E$. $n=1$ 时(30)是显然的: $\bar{u}_i^{(1)}=b_i^{(1)}\leqslant b_i=u_i^{(1)}$. 设(30)对 n 成立, 则

$$\begin{aligned}\sum_{\upsilon=1}^{n+1}\bar{u}_i^{(\upsilon)}&=\bar{u}_i^{(1)}+\sum_{\upsilon=1}^{n}\bar{u}_i^{(\upsilon+1)}=b_i^{(1)}+\sum_{\upsilon=1}^{n}\sum_k a_{ik}\bar{u}_k^{(\upsilon)}+\sum_{\upsilon=1}^{n}b_i^{(\upsilon+1)}\\&=\sum_{\upsilon=1}^{n+1}b_i^{(\upsilon)}+\sum_k a_{ik}\sum_{\upsilon=1}^{n}\bar{u}_k^{(\upsilon)}\leqslant b_i+\sum_k a_{ik}\sum_{\upsilon=1}^{n}u_k^{(\upsilon)}=\sum_{\upsilon=1}^{n+1}u_i^{(\upsilon)}.\end{aligned}\quad\square$$

定理 4.7 对任意的 $j\in E$, $\{m_{ij}, i\in E\}$ 是线性方程组

$$z_i=\sum_{k\neq j}p_{ik}z_k+f_{ij},\quad i\in E\tag{31}$$

的最小非负解.

证 我们已经知道, $n\geqslant 1$ 时 $f_{ij}^{(n+1)}=\sum_{k\neq j}p_{ik}f_{kj}^{(n)}$, 因此

$$(n+1)f_{ij}^{(n+1)}=\sum_{k\neq j}p_{ik}(nf_{kj}^{(n)})+f_{ij}^{(n+1)},$$

而 $\sum_{n=1}^{\infty}f_{ij}^{(n)}=f_{ij}$, $\sum_{n=1}^{\infty}nf_{ij}^{(n)}=m_{ij}$, 由引理 4.7 即知 $\{m_{ij}, i\in E\}$ 为方程组(31)的最小非负解. $\square$

例 4.2 有一个吸收壁的随机游动(续) 我们已经知道, $p\leqslant q$ 时, $f_{j0}=1, j\geqslant 1$. 现在求平均吸收时间 $m_{j0}, j\geqslant 1$. 由定理 4.7, $\{m_{j0}, j\geqslant 1\}$ 是下列方程组的最小非负解:

$$\begin{cases}z_1=pz_2+1,\\ z_j=pz_{j+1}+qz_{j-1}+1,\quad j\geqslant 2.\end{cases}\tag{32}$$

将(32)改写成

$$\begin{cases} z_2 - z_1 = \frac{q}{p} z_1 - \frac{1}{p}, \\ z_{j+1} - z_j = \frac{q}{p}(z_j - z_{j-1}) - \frac{1}{p}, \quad j \geqslant 2, \end{cases}$$

由此可得,$j \geqslant 2$ 时

$$\begin{aligned} z_{j+1} - z_j &= \left(\frac{q}{p}\right)^2 (z_{j-1} - z_{j-2}) - \left[\frac{q}{p} + 1\right]\frac{1}{p} = \cdots \\ &= \left(\frac{q}{p}\right)^{j-1} (z_2 - z_1) - \left[\left(\frac{q}{p}\right)^{j-2} + \cdots + 1\right]\frac{1}{p} \\ &= \left(\frac{q}{p}\right)^{j} z_1 - \left[\left(\frac{q}{p}\right)^{j-1} + \cdots + 1\right]\frac{1}{p}, \\ z_{j+1} - z_1 &= \sum_{v=1}^{j} \left(\frac{q}{p}\right)^{v} z_1 - \sum_{v=1}^{j} \left[\left(\frac{q}{p}\right)^{v-1} + \cdots + 1\right]\frac{1}{p}, \\ z_{j+1} &= \sum_{v=0}^{j} \left(\frac{q}{p}\right)^{v} z_1 - \sum_{v=1}^{j} \left[\left(\frac{q}{p}\right)^{v-1} + \cdots + 1\right]\frac{1}{p}. \end{aligned} \tag{33}$$

先考虑 $p < q$ 的情形.这时

$$z_{j+1} = \sum_{v=0}^{j} \left(\frac{q}{p}\right)^{v} z_1 - \frac{1}{q-p}\left[\sum_{v=1}^{j} \left(\frac{q}{p}\right)^{v} - j\right], \quad j \geqslant 1.$$

我们只讨论非负解:$z_j \geqslant 0, j \geqslant 1$.在上式两端除以 $\sum_{v=0}^{j} (q/p)^v$,并让 $j \to \infty$,即得

$$z_1 \geqslant \frac{1}{q-p},$$

而 $j \geqslant 1$ 时

$$z_{j+1} \geqslant \sum_{v=0}^{j} \left(\frac{q}{p}\right)^{v} \frac{1}{q-p} - \frac{1}{q-p}\left[\sum_{v=1}^{j} \left(\frac{q}{p}\right)^{v} - j\right] = \frac{j+1}{q-p}.$$

另一方面,不难直接验证$\{z_j = \frac{j}{q-p}, j \geqslant 1\}$是方程组(32)的解,因而也是它的最小非负解,即

$$m_{j0} = \frac{j}{q-p}, \quad j \geqslant 1.$$

若 $p = q$,由(33)式应有

$$m_{j+1,0} = (j+1) m_{10} - j(j+1), \quad j \geqslant 2.$$

若 m_{10}为有限数,则 j 充分大之后,m_{j0}将为负数,这是不可能的.因此,只能有 $m_{10} = \infty$,从而对一切 $j \geqslant 1, m_{j0} = \infty$. □

利用定理 4.7 中的方程组(31),只能求为某一吸收状态吸收所需平均时间.如果存在两个或更多的吸收状态,自然希望知道被无论哪一个吸收状态吸收所需平均时间.在赌徒输光问题中,赌博持续的平均时间就是这样的一个例子.下述定理能用来解决这一类问题.

定理 4.8 设 $N=\inf\{n\geqslant 1: X_n \notin T\}$ 为首次进入常返状态的时刻,且对一切 $i\in T$,$\mathrm{P}_i(N<\infty)=1$,即从非常返状态出发迟早要进入常返状态,令 $h_i=E_i[N]$,则 $\{h_i, i\in T\}$ 是线性方程组

$$z_i=\sum_{k\in T}p_{ik}z_k+1,\quad i\in T \tag{34}$$

的最小非负解.

证 记

$$h_i^{(n)}=\mathrm{P}_i(N=n)=\mathrm{P}_i(X_n\notin T, X_v\in T, 0<v<n).$$

由假设,$i\in T$ 时,

$$\sum_{n=1}^{\infty}h_i^{(n)}=\mathrm{P}_i(N<\infty)=1,$$

且 $h_i=\sum\limits_{n=1}^{\infty}nh_i^{(n)}$. $n\geqslant 1$ 时

$$\begin{aligned}
h_i^{(n+1)} &= \sum_{k\in T}\mathrm{P}_i(X_{n+1}\notin T, X_v\in T, 1<v<n+1, X_1=k)\\
&= \sum_{k\in T}p_{ik}\mathrm{P}(X_{n+1}\notin T, X_v\in T, 1<v<n+1\mid X_1=k)=\sum_{k\in T}p_{ik}h_k^{(n)},\\
(n+1)h_i^{(n+1)} &= \sum_{k\in T}p_{ik}(nh_k^{(n)})+h_i^{(n+1)},
\end{aligned}$$

由引理 4.7 即得定理的结论. □

例 4.3 有两个吸收壁的随机游动(续) 我们用定理 4.8 中的记号. 这时 $h_i, 1\leqslant i\leqslant a-1$ 为从状态 i 出发被任一个吸收壁吸收平均所需时间. 在赌徒输光问题中,即为赌徒甲的初始赌本是 i 元时赌博持续的平均时间. 由方程组(34)得

$$\begin{cases}
h_1=ph_2+1,\\
h_i=qh_{i-1}+ph_{i+1}+1, & 2\leqslant i\leqslant a-2,\\
h_{a-1}=qh_{a-2}+1.
\end{cases}$$

令 $h_0=h_a=0$,上述方程组有统一的形式:$1\leqslant i\leqslant a-1$ 时

$$\begin{gathered}
h_i=ph_{i+1}+qh_{i-1}+1,\\
p(h_{i+1}-h_i)=q(h_i-h_{i-1})-1.
\end{gathered}$$

记 $r=q/p$,则 $r\neq 1$,即 $p\neq q$ 时

$$\begin{aligned}
h_{i+1}-h_i &= r(h_i-h_{i-1})-\frac{1}{p}=\cdots\\
&= r^i(h_1-h_0)-(r^{i-1}+\cdots+1)\frac{1}{p}\\
&= r^i(h_1-h_0)+\frac{r^i-1}{p-q},\\
h_{i+1}-h_1 &= \sum_{k=1}^{i}r^k(h_1-h_0)+\frac{1}{p-q}\left(\sum_{k=1}^{i}r^k-i\right),
\end{aligned}$$

$$h_{i+1}=\frac{r^{i+1}-1}{r-1}h_1+\frac{1}{p-q}\left(\frac{r^{i+1}-r}{r-1}-i\right),\quad 1\leqslant i\leqslant a-1.$$

由 $h_a=0$ 可求得 h_1:

$$0=\frac{r^a-1}{r-1}h_1+\frac{1}{p-q}\left[\frac{r^a-r}{r-1}-(a-1)\right],$$

$$h_1=\frac{1}{p-q}\left(a\,\frac{r-1}{r^a-1}-1\right),$$

$$h_i=\frac{r^i-1}{r-1}h_1+\frac{1}{p-q}\left[\frac{r^i-r}{r-1}-(i-1)\right]$$

$$=\frac{1}{p-q}\left(a\,\frac{r^i-1}{r^a-1}-i\right),\quad 1\leqslant i\leqslant a-1.$$

在 $r=1$,即 $p=q=1/2$ 时,可作类似的计算:对 $0\leqslant i\leqslant a-1$,

$$h_{i+1}-h_i=(h_i-h_{i-1})-2=\cdots=(h_1-h_0)-2i,$$

$$h_{i+1}=(i+1)h_1-i(i+1).$$

而

$$0=h_a=ah_1-a(a-1),\qquad h_1=a-1,$$

因此,$1\leqslant i\leqslant a-1$ 时

$$h_i=ih_1-i(i-1)=i(a-1)-i(i-1)=i(a-1).\quad\square$$

习　题

4－1. 设转移概率矩阵为:

(1) $\begin{pmatrix}1/2 & 1/4 & 1/4\\ 0 & 1/2 & 1/2\\ 0 & 1 & 0\end{pmatrix}$;　(2) $\begin{pmatrix}1 & 0 & 0\\ 1/2 & 1/4 & 1/4\\ 0 & 0 & 1\end{pmatrix}$;

(3) $\begin{pmatrix}1 & 0 & 0 & 0\\ 2/3 & 0 & 1/3 & 0\\ 0 & 2/3 & 0 & 1/3\\ 0 & 0 & 0 & 1\end{pmatrix}$;　(4) $\begin{pmatrix}1 & 0 & 0 & 0\\ 2/5 & 2/5 & 1/5 & 0\\ 0 & 1/5 & 2/5 & 2/5\\ 0 & 0 & 0 & 1\end{pmatrix}$;

(5) $\begin{pmatrix}1 & 0 & 0 & 0\\ 1/4 & 1/4 & 1/4 & 1/4\\ 0 & 0 & 1/2 & 1/2\\ 0 & 0 & 1 & 0\end{pmatrix}$;　(6) $\begin{pmatrix}0 & 1/2 & 1/2 & 0\\ 1/2 & 0 & 0 & 1/2\\ 0 & 0 & 1/3 & 2/3\\ 0 & 0 & 1/2 & 1/2\end{pmatrix}$,

求矩阵(f_{ij}).

4－2. 设转移概率为

$$p_{00}=1,p_{10}=(1-\delta)q,p_{11}=\delta q,p_{12}=p,$$

$$p_{ii}=q,\quad p_{i,i+1}=p,\quad i\geqslant 2,$$

$$0<p,\delta<1,\quad p+q=1.$$

求吸收概率 $f_{i0}, i\geqslant 1$.

4-3. 设转移概率为

$$p_{00}=1, p_{10}=(1-\delta)q, p_{11}=\delta q, p_{12}=p,$$
$$p_{i,i-1}=q,\ p_{i,i+1}=p,\quad i\geqslant 2,$$
$$0<p,\delta<1,\quad p+q=1.$$

求吸收概率 $f_{i0}, i\geqslant 1$.

4-4. 设状态空间为 $E=\{0,1,\cdots,a\}$,转移概率满足条件:

$$\sum_{j=0}^{a} jp_{ij}=i,\quad i=0,\cdots,a;\quad p_{ii}<1,\quad i=1,\cdots,a-1.$$

证明:0 及 a 为吸收状态,且

$$f_{i0}=1-\frac{i}{a},\quad f_{ia}=\frac{i}{a},\quad i=1,\cdots,a-1.$$

4-5. 设状态空间为 $E=\{0,1,\cdots,a\}$.考虑下列两个转移概率矩阵:

(1) $p_{ij}=\binom{a}{j}\left(\frac{i}{a}\right)^{j}\left(1-\frac{i}{a}\right)^{a-j},\quad 0\leqslant i,j\leqslant a;$

(2) $p_{00}=p_{aa}=1,\quad p_{ii}=1-2\,\frac{i(a-i)}{a^{2}},\quad 0<i<a,$

$p_{ij}=\frac{i(a-i)}{a^{2}},\quad |j-i|=1,\quad 0<i<a,$

$p_{ij}=0,\quad |j-i|\geqslant 2.$

在这两个转移概率矩阵之下,0 与 a 都是吸收状态.证明:在这两个转移概率矩阵之下,吸收概率 $f_{i0}, f_{ia}, 0<i<a$ 是相同的.

4-6. 在马尔可夫链中定义

$$P_{ij}(z)=\sum_{n=0}^{\infty} p_{ij}^{(n)}z^{n},\ F_{ij}(z)=\sum_{n=0}^{\infty} f_{ij}^{(n)}z^{n},\ |z|<1.$$

证明:对一切 $i,j\in E$

$$P_{ij}(z)=\delta_{ij}+F_{ij}(z)P_{jj}(z),\quad |z|<1.$$

由此可得

$$P_{ii}(z)=\frac{1}{1-F_{ii}(z)},\quad |z|<1.$$

令 $z\to 1$,由于 $F_{ii}(1)=f_{ii}$,我们又得到常返判别法:$f_{ii}=1$ 等价于 $\sum_{n=0}^{\infty} p_{ii}^{(n)}=\infty$ 的一个证明.

4-7. 在(p,q)随机游动中,证明:

$$f_{00}^{(2n)}=\begin{cases}2pq, & n=1,\\ \frac{1}{2}\frac{3}{2}\cdots\frac{2n-3}{2}\frac{(4pq)^n}{2(n!)}, & n\geqslant 2.\end{cases}$$

4－8. 在(p,q)随机游动中，记$\xi=\sup_{n\geqslant 0}X_n$. 证明：

(1) $p\geqslant q$ 时，$\mathrm{P}(\xi=\infty\mid X_0=0)=1$；

(2) $p<q$ 时，

$$\mathrm{P}(\xi=n\mid X_0=0)=\left(\frac{p}{q}\right)^n\left(1-\frac{p}{q}\right),\quad n=0,1,\cdots.$$

4－9. 证明：在(p,q)随机游动中，对一切 $i\in\mathbb{Z}$，

$$f_{ii}=1-|p-q|.$$

4－10. 设马尔可夫链没有吸收状态. 定义

$$m_0=0,$$
$$m_n=\inf\{n>m_{n-1},Xn\neq X_{m_{n-1}}\},\ n\geqslant 1.$$

(由于没有吸收状态，在任一状态上停留的时间是有限的，因此 m_n 是有限数.)令

$$Y_n=X_{m_n},\qquad n\geqslant 0.$$

证明：(1)　$Y=\{Y_n,n\geqslant 0\}$为马尔可夫链，其转移概率矩阵为

$$\left((1-\delta_{ij})\frac{p_{ij}}{1-p_{ii}}\right);$$

(2) Y 为常返不可约的充要条件是$X=\{X_n,n\geqslant 0\}$为常返不可约的；

(3) 对任意的 $n\geqslant 0$

$$\mathrm{P}(m_{n+1}-m_n=k\mid Y_n=i)=(1-p_{ii})p_{ii}^{k-1},\quad k\geqslant 1,$$

且对任意的 $n\geqslant 1$

$$\mathrm{P}\Big(\bigcap_{j=1}^{n}\{m_j-m_{j-1}=k_j\}\mid\bigcap_{j=0}^{n-1}\{Y_j=i_j\}\Big)=\prod_{j=0}^{n-1}\big[(1-p_{i_ji_j})p_{i_ji_j}^{k_{j+1}-1}\big].$$

这表明，在已知 Y 的条件下，$\{m_n-m_{n-1},n\geqslant 1\}$是独立随机变量序列，且每一个变量服从几何分布. 事实上，原来的马尔可夫链 X 可以反过来用 Y 及$\{m_n,n\geqslant 1\}$构造出来：

$$X_n=\begin{cases}Y_0, & n<m_1,\\ Y_k, & m_k\leqslant n<m_{k+1},\ k\geqslant 1.\end{cases}$$

4－11. 设 $b_i\geqslant 0,i\in T$. 证明：

(1) 线性方程组

$$z_i=\sum_{k\in T}p_{ik}z_k+b_i,\quad i\in T$$

的最小非负解为

$$z_i^* = \sum_{k\in T}\Big(\sum_{n=0}^{\infty} p_{ik}^{(n)}\Big)b_k, \qquad i\in T;$$

(2) 线性方程组

$$z_i = \sum_{k\in T} z_k p_{ki} + b_i, \qquad i\in T$$

的最小非负解为

$$z_i^* = \sum_{k\in T} b_k\Big(\sum_{n=0}^{\infty} p_{ki}^{(n)}\Big), \qquad i\in T.$$

4-12. 设 $n_{ij} = \sum_{n=0}^{\infty} p_{ij}^{(n)}$, $\boldsymbol{N} = (n_{ij})_{i,j\in T}$, $\boldsymbol{Q} = (p_{ij})_{i,j\in T}$. 证明：

(1) $\boldsymbol{Q}^n = (p_{ij}^{(n)})_{i,j\in T}$;

(2) $\boldsymbol{N}$ 是矩阵方程

$$\boldsymbol{Z} = \boldsymbol{QZ} + \boldsymbol{I}$$

或

$$\boldsymbol{Z} = \boldsymbol{ZQ} + \boldsymbol{I}$$

的最小非负解，特别，当链为有限时，$\boldsymbol{N} = (\boldsymbol{I} - \boldsymbol{Q})^{-1}$.

4-13. 设链的常返状态全都为吸收状态. 证明：$i\in T, j\notin T$ 时

(1) $f_{ij} = \sum_{k\in T}\left(\sum_{n=0}^{\infty} p_{ik}^{(n)}\right)p_{kj}$;

(2) $m_{ij} = \sum_{k\in T}\left(\sum_{n=0}^{\infty} p_{ik}^{(n)}\right)f_{kj}$.

4-14. 设马尔可夫链为非常返不可约的. 固定 $j\in E$，令

$$h_i = \sum_{n=0}^{\infty} p_{ij}^{(n)}, \quad i\in E.$$

证明：$\{h_i, i\in E\}$ 为非常数的过分函数.

4-15. 设在分支过程中每一个体至多产生两个后代. 证明：以概率 1 灭种的充要条件是 $g_0 \geq g_2$.

4-16. 设某种生物的每一个有生殖力的个体产生三个后代，每个后代有无生殖力是等可能的，无生殖力的个体随即死亡，求灭种概率.

4-17. 设马尔可夫链有三个状态，且

$$(m_{ij}) = \begin{pmatrix} 2 & 3 & 3 \\ 1 & 4 & 4 \\ 1 & 4 & 4 \end{pmatrix}.$$

求转移概率矩阵.

4-18. 设马尔可夫链有三个状态，且

$$(m_{ij})=\begin{pmatrix}a&2&6\\4&a&4\\6&2&a\end{pmatrix}.$$

求常数 a.

4－19. 若一篇文稿有 n 个错误.每校阅一次至少能发现一个错误,但遗留下来的错误个数在 0 到 $n-1$ 之间是等可能的.设原稿共有 a 个错误.求为了改正全部错误,平均需要校阅几次?

4－20. 设有 a 个选手依次两两进行比赛,获胜者接连比赛下去(擂台赛).比赛进行到有一名选手连续地击败其余 $a-1$ 名选手为止.假设每场比赛中,两名选手获胜的概率各为 1/2.求平均需比赛多少次?

§2.5 平稳分布

定义 不恒为零的非负数列 $\{u_i, i\in E\}$ 称为马尔可夫链的不变测度,若对一切 $i\in E$,

$$u_i=\sum_k u_k p_{ki}. \tag{1}$$

若不变测度 $\{u_i, i\in E\}$ 是一个概率分布列,即 $\sum_i u_i=1$,则称它为马尔可夫链的平稳分布.

若 $\{u_i, i\in E\}$ 为不变测度,则对一切 $n\geqslant1$,

$$u_i=\sum_k u_k p_{ki}^{(n)},\quad i\in E. \tag{2}$$

(2)式可用归纳法证明. $n=1$ 时(2)式即(1)式.若(2)式对 n 成立,则它对 $n+1$ 也成立:

$$u_i=\sum_k u_k p_{ki}^{(n)}=\sum_k\sum_j u_j p_{jk} p_{ki}^{(n)}=\sum_j u_j\sum_k p_{jk} p_{ki}^{(n)}=\sum_j u_j p_{ji}^{(n+1)}.$$

若 $\{u_i, i\in E\}$ 为不变测度,且 $u=\sum_i u_i<\infty$,则易见,规范化后得到的 $\{u_i/u, i\in E\}$ 是平稳分布.

如果 $\{u_i, i\in E\}$ 是一个平稳分布,把它取为初始分布:

$$\mathsf{P}(X_0=i)=u_i,\qquad i\in E,$$

则由(2)式,对一切 $n\geqslant1$

$$\mathsf{P}(X_n=i)=\sum_k \mathsf{P}(X_0=k)\,p_{ki}^{(n)}=\sum_k u_k p_{ki}^{(n)}=u_i,\quad i\in E.$$

更进一步,对任意的 $t_1<t_2<\cdots<t_n$ 及 $m\geqslant1$, $(X_{t_1},\cdots,X_{t_n})$ 的分布与 $(X_{t_1+m},\cdots,X_{t_n+m})$ 的分布完全相同:

$$\begin{aligned}\mathsf{P}(X_{t_1}=i_1,\cdots,X_{t_n}=i_n)&=u_{i_1}p_{i_1i_2}^{(t_2-t_1)}\cdots p_{i_{n-1}i_n}^{(t_n-t_{n-1})}\\&=\mathsf{P}(X_{t_1+m}=i_1,\cdots,X_{t_n+m}=i_n),\end{aligned}$$

即有限维分布经过时间的推移保持不变: $\{X_n, n\geqslant0\}$ 是平稳过程.平稳分布的名称即由此而来.

例 5.1 对 (p,q) 随机游动,易见 $\{u_i\equiv1, i\in\mathbb{Z}\}$ 是不变测度.若 $p\neq q$,则 $\{u_i=(p/q)^i, i\in\mathbb{Z}\}$ 也是不变测度:

$$\sum_k u_k p_{ki} = \left(\frac{p}{q}\right)^{i-1} p + \left(\frac{p}{q}\right)^{i+1} q = \left(\frac{p}{q}\right)^{i} (q+p) = \left(\frac{p}{q}\right)^{i} = u_i . \quad \square$$

对任意的 $i, j \in E$,令

$$e_{ij}^{(n)} = \mathrm{P}_i(X_n = j, X_v \neq i, 0 < v < n) \tag{3}$$

为从状态 i 出发经过 n 步到达状态 j 但中途不经过 i 的概率. 易见 $e_{ij}^{(1)} = p_{ij}, e_{ii}^{(n)} = f_{ii}^{(n)}, n \geq 1$. 令

$$e_{ij} = \sum_{n=1}^{\infty} e_{ij}^{(n)}. \tag{4}$$

我们有 $e_{ii} = f_{ii}$,且 $i \not\to j$ 时 $e_{ij} = 0$. 但 $i \to j$ 时一般并不知道 $e_{ij} < \infty$.

定理 5.1 设 j 为常返状态,则 $\{u_i = e_{ji}, i \in E\}$ 为不变测度.

证 由定义(3),

$$\begin{aligned} e_{ji}^{(1)} &= p_{ji}, \\ e_{ji}^{(n+1)} &= \mathrm{P}_j(X_{n+1} = i, X_v \neq j, 0 < v < n+1) \\ &= \sum_{k \neq j} \mathrm{P}_j(X_{n+1} = i, X_n = k, X_v \neq j, 0 < v < n) \\ &= \sum_{k \neq j} \mathrm{P}_j(X_n = k, X_v \neq j, 0 < v < n) \mathrm{P}(X_{n+1} = i \mid X_n = k) \\ &= \sum_{k \neq j} e_{jk}^{(n)} p_{ki}, \quad n \geq 1. \end{aligned}$$

由引理 4.1,$\{u_i = e_{ji}, i \in E\}$ 是线性方程组

$$z_i = \sum_{k \neq j} z_k p_{ki} + p_{ji}, \quad i \in E \tag{5}$$

的最小非负解. 注意到 $u_j = e_{jj} = 1$,有

$$u_i = \sum_{k \neq j} u_k p_{ki} + p_{ji} = \sum_k u_k p_{ki}, \quad i \in E.$$

显然,$\{u_i, i \in E\}$ 是不恒为零的、非负的,故 $\{u_i, i \in E\}$ 满足(1),从而满足(2). 还需说明所有的 $u_i = e_{ji} < \infty$. 若 $j \not\to i$,我们早已知道 $u_i = e_{ji} = 0$. 若 $j \to i$,因 j 是常返状态,也有 $i \to j$,所以有 $n \geq 1$ 使 $p_{ij}^{(n)} > 0$. 因此

$$1 = u_j = \sum_k u_k p_{kj}^{(n)} \geq u_i p_{ij}^{(n)}, \quad u_i \leq 1/p_{ij}^{(n)} < \infty. \qquad \square$$

定理 5.2 设链为常返不可约的,则对一切 $i, j \in E$

$$\lim_{N \to \infty} \frac{\sum_{n=0}^{N} p_{ij}^{(n)}}{\sum_{n=0}^{N} p_{ii}^{(n)}} = e_{ij} < \infty. \tag{6}$$

证 $i = j$ 时,$e_{ii} = 1$,(6)是显然的. 所以只需讨论 $i \neq j$ 的情形. 由定理 5.1 已知 $e_{ij} < \infty$. 我们有

$$p_{ij}^{(n)}=\mathrm{P}_i(X_n=j)=\sum_{\upsilon=0}^{n-1}\mathrm{P}_i(X_n=j,X_\upsilon=i,X_l\neq i,\upsilon<l<n)$$
$$=\sum_{\upsilon=0}^{n-1}\mathrm{P}_i(X_\upsilon=i)\mathrm{P}(X_n=j,X_l\neq i,\upsilon<l<n\mid X_\upsilon=i)$$
$$=\sum_{\upsilon=0}^{n-1}p_{ii}^{(\upsilon)}e_{ij}^{(n-\upsilon)}. \tag{7}$$

注意,这里我们是按在时刻 n 之前最后一次进入状态 i 的时刻作事件的分解的,它恰与首次进入法相对.由(7)式推得(6)式的过程,与比极限定理 3.5 中的证明过程完全一样(这并不奇怪,因为我们现在证明的也是一种比极限定理).先求和,

$$\sum_{n=0}^{N}p_{ij}^{(n)}=\sum_{n=1}^{N}\sum_{\upsilon=0}^{n-1}p_{ii}^{(\upsilon)}e_{ij}^{(n-\upsilon)}=\sum_{\upsilon=0}^{N-1}p_{ii}^{(\upsilon)}\sum_{n=1}^{N-\upsilon}e_{ij}^{(n)},$$

再用引理 3.1,即得(6)式. □

引理 5.1 设链有不变测度 $\{u_i,i\in E\}$.若 $u_i>0,i\to j$,则 $u_j>0$.特别,若链又是不可约的,则一切 $u_j>0$.

证 这时有 $n\geqslant1$ 使 $p_{ij}^{(n)}>0$,从而

$$u_j=\sum_k u_kp_{kj}^{(n)}\geqslant u_ip_{ij}^{(n)}>0.$$

不变测度不恒为零,至少有一个 $u_i>0$.若链又是不可约的,则对一切状态 $j,i\to j$,从而 $u_j>0$.
□

定理 5.3 设链为常返不可约的,则不计一个常数因子,不变测度是唯一的.

证 取定一个状态 j.由定理 5.1,$\{e_{ji},i\in E\}$是一个不变测度.现在设$\{u_i,i\in E\}$为不变测度.由引理 5.1,可定义

$$q_{ik}=\frac{u_k}{u_i}p_{ki},\quad i,k\in E,$$

则

$$q_{ik}\geqslant0,\ \sum_k q_{ik}=\frac{1}{u_i}\sum_k u_kp_{ki}=\frac{1}{u_i}u_i=1,$$

因此(q_{ik})可看作一个马尔可夫链的转移概率矩阵,它的 n 步转移概率

$$q_{ik}^{(n)}=\frac{u_k}{u_i}p_{ki}^{(n)},\quad i,k\in E.$$

这可用归纳法容易得到.由此可见,这个链也是不可约的,且是常返的,因为 $\sum_n q_{ii}^{(n)}=\sum_n p_{ii}^{(n)}=\infty$.由定理 3.5 及定理 5.2,在下式中取极限:

$$\frac{\sum_{n=0}^{N}q_{ij}^{(n)}}{\sum_{n=0}^{N}q_{jj}^{(n)}}=\frac{u_j}{u_i}\frac{\sum_{n=0}^{N}p_{ji}^{(n)}}{\sum_{n=0}^{N}p_{jj}^{(n)}},$$

我们得到

$$1=\frac{u_j}{u_i}e_{ji},\quad u_i=u_je_{ji},\quad i\in E. \tag{8}$$

即$\{u_i,i\in E\}$与$\{e_{ji},i\in E\}$只差一个常数. □

推论 设链为常返不可约的,则对任意的$i,j,k\in E$,

$$e_{ij}e_{jk}=e_{ik}. \tag{9}$$

证 $\{e_{ik},k\in E\}$也为不变测度.由(8)式即得(9)式. □

引理 5.2 设$\{u_i,i\in E\}$为一平稳分布,$u_j>0$,则j是常返状态.

证 反证之.设j为非常返状态,则对一切i,$\lim_{n\to\infty}p_{ij}^{(n)}=0$.由于$\sum_i u_i=1$,

$$u_j=\sum_i u_ip_{ij}^{(n)}$$

右边的级数关于n一致收敛,因此令$n\to\infty$即得$u_j=0$,与假设矛盾.故j为常返状态. □

引理 5.3

$$\sum_i e_{ji}=\begin{cases}m_{jj}, & 若\ j\ 为常返状态,\\ \infty, & 若\ j\ 为非常返状态.\end{cases} \tag{10}$$

证 对任意的$n\geqslant 1$及$i,j\in E$

$$\begin{aligned}\sum_i e_{ji}^{(n)}&=\sum_i \mathrm{P}_j(X_n=i,X_v\neq j,0<v<n)\\&=\mathrm{P}_j(X_v\neq j,0<v<n)=\mathrm{P}_j(r_j\geqslant n)\\&=(1-f_{jj})+\sum_{v=n}^{\infty}f_{jj}^{(v)},\end{aligned}$$

$$\sum_i e_{ji}=\sum_i\sum_{n=1}^{\infty}e_{ji}^{(n)}=\sum_{n=1}^{\infty}\left[(1-f_{jj})+\sum_{v=n}^{\infty}f_{jj}^{(v)}\right].$$

因此,若j为非常返状态,$f_{jj}<1$,则$\sum_i e_{ji}=\infty$;若j为常返状态,$f_{jj}=1$,则

$$\sum_i e_{ji}=\sum_{n=1}^{\infty}\sum_{v=n}^{\infty}f_{jj}^{(v)}=\sum_{v=1}^{\infty}vf_{jj}^{(v)}=m_{jj}. \qquad □$$

定义 设i为常返状态,若$m_{ii}<\infty$,称i为正常返状态;若$m_{ii}=\infty$,称i为零常返状态.非周期正常返状态称为遍历状态.

正常返及零常返状态的概率意义是明显的:虽然i是常返状态,从i出发必定要返回i,但平均返回时间可能是有限的或无限的,也就是说返回的速度有快有慢.将常返状态分为正的及零的,正是在这点上作出区别.今后记

$$\pi_i=\begin{cases}\dfrac{1}{m_{ii}}, & 若\ i\ 为正常返状态,\\ 0, & 其它情形.\end{cases} \tag{11}$$

定理 5.4 设链为不可约的.若链有平稳分布,则全部状态是正常返的.反之,若有一个状态

是正常返的，则链有唯一的平稳分布：$\{\pi_i, i \in E\}$，且对一切 $i, j \in E, e_{ij} = \pi_j / \pi_i$.

证 设不可约链有平稳分布 $\{u_i, i \in E\}$，则由引理 5.1 及 5.2 可知，链是常返的，且由定理 5.3，平稳分布是唯一的. 任取状态 j，则 $\{e_{ji}, i \in E\}$ 是不变测度. 由不变测度的唯一性及引理 5.3，

$$m_{jj} = \sum_i e_{ji} < \infty.$$

因此 j 是正常返的，且

$$u_i = \frac{e_{ji}}{m_{jj}}, i \in E. \tag{12}$$

特别，

$$u_j = \frac{1}{m_{jj}} = \pi_j. \tag{13}$$

(13) 对一切状态 j 成立. 再由(12)式，对一切 $i \in E$，

$$\pi_i = \frac{e_{ji}}{m_{jj}} = e_{ji}\pi_j.$$

反之，若有一个状态 j 是正常返的，则如上所知，$\{e_{ji}/m_{jj}, i \in E\}$ 是平稳分布. □

由定理 5.4 可知，同一个常返类中的状态或全为正常返的，或全为零常返的. 相应地，我们就称这个类为*正常返类*或*零常返类*. 因此，定理 5.4 表明，一个不可约链有平稳分布的充要条件是链为正常返的. 实际上，这提供了判别一个不可约链为正常返的最有用的准则：考察线性方程组

$$z_i = \sum_k z_k p_{ki}, \quad i \in E \tag{14}$$

是否有收敛的非零非负解. 如果没有的话，链是非常返的；如果有的话，链是正常返的，再求出这个解，就可得到平稳分布. 需注意的是，(14)的解允许一个任意的常数因子. 但平稳分布是概率分布列，它必须满足总和等于 1 这个规范化条件.

推论 不可约有限马尔可夫链必是正常返的，有唯一的平稳分布.

证 不可约有限链必是常返的. 故对任一状态 j，$\{e_{ji}, i \in E\}$ 是不变测度. 但由于状态空间有限，$m_{jj} = \sum_i e_{ji} < \infty$. 因此，链是正常返的. □

例 5.2 设转移概率矩阵为

$$\mathbf{P} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/4 & 3/4 \\ 1 & 0 & 0 \end{pmatrix}.$$

易见，这是一个不可约有限链. 它的平稳分布 $\{\pi_0, \pi_1, \pi_2\}$ 满足方程组

$$\begin{cases} \pi_0 = \pi_2, \\ \pi_1 = \dfrac{1}{2}\pi_0 + \dfrac{1}{4}\pi_1, \\ \pi_2 = \dfrac{1}{2}\pi_0 + \dfrac{3}{4}\pi_1. \end{cases}$$

这里实际上只有两个独立的方程,解得:

$$\pi_1=\frac{2}{3}\pi_0,\qquad \pi_2=\pi_0.$$

再利用规范化条件 $\pi_0+\pi_1+\pi_2=1$ 得

$$\left(1+\frac{2}{3}+1\right)\pi_0=1,$$

$$\pi_0=\frac{3}{8},\quad \pi_1=\frac{1}{4},\quad \pi_2=\frac{3}{8}.\qquad \square$$

例 5.3 设灯泡的寿命以天计算,一个灯泡使用到第 k 天坏了,它的寿命就是 k. 一个灯泡坏了,第二天再换上一个. 以 $\{\xi_n,n\geqslant1\}$ 记相继换上的一列灯泡的寿命. 设 $\{\xi_n,n\geqslant1\}$ 为独立同分布随机变量,其共同分布为

$$\mathrm{P}(\xi_n=k)=a_k,\qquad k\geqslant1.$$

对 $n\geqslant0$ 定义

$$Y_n=\begin{cases}0, & n\leqslant\xi_1,\\ k, & \xi_1+\cdots+\xi_k<n\leqslant\xi_1+\cdots+\xi_{k+1},\quad k\geqslant1,\end{cases}$$

$$X_n=n-\sum_{j=0}^{Y_n}\xi_j\qquad(\xi_0=0),$$

X_n 是第 n 天正在使用的灯泡已经使用的天数(当天计算在内). 这时 $\{X_n,n\geqslant0\}$ 是状态空间为 $\{1,2,\cdots\}$ 的马尔可夫链,其转移概率矩阵为

$$\mathbf{P}=\begin{pmatrix} a_1 & 1-a_1 & 0 & 0 & \cdots\\ \frac{a_2}{1-a_1} & 0 & 1-\frac{a_2}{1-a_1} & 0 & \cdots\\ \frac{a_3}{1-a_1-a_2} & 0 & 0 & 1-\frac{a_3}{1-a_1-a_2} & \cdots\\ \cdots & \cdots & \cdots & \cdots & \cdots\end{pmatrix}.$$

显然这是一个不可约链,也不难看出,$f_{11}^{(1)}=a_1$, $n\geqslant2$ 时,

$$f_{11}^{(n)}=(1-a_1)\left(1-\frac{a_2}{1-a_1}\right)\cdots\left(1-\frac{a_{n-1}}{1-a_1-\cdots-a_{n-2}}\right)\frac{a_n}{1-a_1-\cdots-a_{n-1}}=a_n.$$

因此,$f_{11}=\sum_{n=1}^{\infty}f_{11}^{(n)}=\sum_{n=1}^{\infty}a_n=1$. 链是常返的. 链是正常返的充要条件为 $\sum_{k=1}^{\infty}ka_k<\infty$. $\square$

例 5.4 有一个反射壁的随机游动 我们继续例4.5的讨论. 记

$$p_k=\frac{q_1\cdots q_k}{p_1\cdots p_k},\quad k\geqslant1.$$

前面我们已经证明了,链为常返的充要条件是 $\sum_k \rho_k=\infty$. 继续讨论正常返的条件. 为此讨论线性方程组

$$
\begin{cases} z_0=r_0z_0+q_1z_1, \\ z_n=p_{n-1}z_{n-1}+r_nz_n+q_{n+1}z_{n+1}, \quad n\geqslant 1. \end{cases} \tag{15}
$$

注意到 $1-r_n=p_n+q_n$,方程组可化成

$$
\begin{cases} q_1z_1-p_0z_0=0, \\ q_{n+1}z_{n+1}-p_nz_n=q_nz_n-p_{n-1}z_{n-1}, \quad n\geqslant 1. \end{cases}
$$

由此可知

$$
q_{n+1}z_{n+1}-p_nz_n=0, \qquad n\geqslant 0,
$$

$$
z_{n+1}=\frac{p_n}{q_{n+1}}z_n=\cdots=\frac{p_n\cdots p_0}{q_{n+1}\cdots q_1}z_0=\frac{p_0}{p_{n+1}}\frac{1}{\rho_{n+1}}z_0.
$$

因此,方程组(15)有收敛解,即链为正常返的充要条件为

$$
\sum_{n=1}^{\infty}\frac{1}{p_n\rho_n}<\infty
$$

(这条件蕴含了 $\sum_n \rho_n=\infty$),且这时平稳分布为

$$
\pi_0=\left(1+p_0\sum_{n=1}^{\infty}\frac{1}{p_n\rho_n}\right)^{-1}, \quad \pi_n=\frac{p_0}{p_n\rho_n}\pi_0, \quad n\geqslant 1.
$$

在 $p_i=p, q_i=q, i\geqslant 1$ 的特殊情形,

$$
\sum_{n=1}^{\infty}\frac{1}{p_n\rho_n}=\frac{1}{p}\sum_{n=1}^{\infty}\left(\frac{p}{q}\right)^n.
$$

因此链为正常返的充要条件 $p<q$,且这时平稳分布为

$$
\pi_0=\frac{q-p}{q-p+p_0}, \quad \pi_n=\frac{p_0(q-p)}{p(q-p+p_0)}\left(\frac{p}{q}\right)^n, \quad n\geqslant 1.
$$

若又有 $p_0=p$,则平稳分布是几何分布:

$$
\pi_n=\left(1-\frac{p}{q}\right)\left(\frac{p}{q}\right)^n, \qquad n\geqslant 0.
$$

把所得结果总结起来:

(1) 链为非常返 $\Leftrightarrow \sum_{n=1}^{\infty}\rho_n<\infty$;

(2) 链为零常返 $\Leftrightarrow \sum_{n=1}^{\infty}\rho_n=\infty$, $\sum_{n=1}^{\infty}\frac{1}{p_n\rho_n}=\infty$;

(3) 链为正常返 $\Leftrightarrow \sum_{n=1}^{\infty}\frac{1}{p_n\rho_n}<\infty$.

当 $p=p_i, q_i=q, i\geqslant 1$ 时,

(1) 链为非常返$\Leftrightarrow$ $p>q$;

(2) 链为零常返$\Leftrightarrow$ $p=q$;

(3) 链为正常返$\Leftrightarrow$ $p<q$. □

定理 5.5 设链为常返不可约的,则线性不等式组

$$z_i \geqslant \sum_k z_k p_{ki}, \quad i \in E$$

的不恒为零的非负解$\{u_i, i \in E\}$必是不变测度.

证 任取状态 j,定义

$$q_{ik} = \frac{e_{jk}}{e_{ji}} p_{ki}, \quad i, k \in E,$$

则以(q_{ik})为转移概率矩阵的马尔可夫链也是常返不可约的,且$\{u_i/e_{ji}, i \in E\}$是它的过分函数:

$$\sum_k q_{ik} \frac{u_k}{e_{jk}} = \frac{1}{e_{ji}} \sum_k u_k p_{ki} \leqslant \frac{u_i}{e_{ji}}, \quad i \in E.$$

由定理 4.6,常返不可约链的过分函数必为常数,而$\{e_{ji}, i \in E\}$是不变测度,所以$\{u_i, i \in E\}$也是不变测度. □

定理 5.6 不可约链为正常返的充要条件是线性不等式组

$$z_i \geqslant \sum_k z_k p_{ki}, \quad i \in E \tag{16}$$

的非零非负解均为收敛解.

证 先证必要性.由定理 5.5,线性不等式组(16)的非零非负解$\{u_i, i \in E\}$必是不变测度,且 $u_i = ce_{ji}, i \in E$,其中 c 为常数,j 为任一取定的状态.因此,

$$\sum_i u_i = c \sum_i e_{ji} = cm_{jj} < \infty.$$

再证充分性.任取状态 j,令 $u_i = e_{ji}$,则$\{u_i, i \in E\}$满足不等式组(16).事实上,在定理 5.1 的证明中,我们已指出$\{u_i, i \in E\}$满足方程组

$$z_i = \sum_{k \neq j} z_k p_{ki} + p_{ji}, \quad i \in E,$$

而 $u_j = e_{jj} = f_{jj} \leqslant 1$,

$$u_i = \sum_{k \neq j} u_k p_{ki} + p_{ji} \geqslant \sum_k u_k p_{ki}, \quad i \in E,$$

这时 $\sum_i u_i = \sum_i e_{ji} < \infty$,由引理 5.3,$j$ 为正常返状态. □

在定理 5.3 及 5.5 的证明中,我们利用了一个很有用的技巧:若$\{u_i, i \in E\}$是一个不变测度,则 $q_{ij} = \frac{u_j}{u_i} p_{ji}, i, j \in E$ 也是一个马尔可夫链的转移概率.它把转移概率 p_{ij} 中 i 与 j 的地位相互交换,即把出发状态与到达状态相互交换,也就是把时间次序颠倒过来.事实上,若$\{u_i, i \in E\}$是平稳分布,且被取作为初始分布,那么把时间颠倒过来而得到的马尔可夫链$\{X_n, X_{n-1}, \cdots,$

$X_1,X_0\}$的转移概率正是 q_{ij}. 由此导出下面的概念.

定义　若存在不恒为零的非负数列$\{u_i,i\in E\}$使得

$$u_ip_{ij}=u_jp_{ji},\qquad i,j\in E, \tag{17}$$

则称马尔可夫链为可配称的. 这时$\{u_i,i\in E\}$必为不变测度：

$$\sum_i u_ip_{ij}=\sum_i u_jp_{ji}=u_j.$$

若满足(17)的$\{u_i,i\in E\}$是一概率分布,则称马尔可夫链为可逆的. 这时$\{u_i,i\in E\}$必为平稳分布. 若以此平稳分布为初始分布,则对一切 n,$(X_0,X_1,\cdots,X_n)$与$(X_n,X_{n-1},\cdots,X_0)$同分布. 这正是可逆的意义. 显然,可逆不可约链必是正常返的.

例 5.4　有一个反射壁的随机游动(续)　这时(17)化为

$$u_0p_0=u_1q_1,\qquad u_ip_i=u_{i+1}q_{i+1},\qquad i\geqslant 1.$$

只需取

$$u_0=1,\quad u_i=\frac{p_{i-1}\cdots p_0}{q_i\cdots q_1},\quad i\geqslant 1$$

就能满足它. 因此链是可配称的. 当链是正常返时,链是可逆的. □

例 5.5　马尔可夫链称为对称的,若

$$p_{ij}=p_{ji},\quad i,j\in E.$$

这时链是可配称的. 为此,只需在(16)中取 $u_i\equiv 1$. 显然,对称马尔可夫链是可逆的当且仅当链是有限状态的. □

例 5.6　设转移概率为(例 1.3)

$$p_{ij}=\mathrm{e}^{-\lambda}\sum_{k=0}^{\min(i,j)}\mathrm{C}_i^k(1-q)^kq^{i-k}\frac{\lambda^{j-k}}{(j-k)!},\quad \lambda>0,0<q<1.$$

由于对一切 $i,j\in E$, $p_{ij}>0$,链是非周期不可约的. 令 $u_i=\frac{1}{i!}\left(\frac{\lambda}{q}\right)^i\mathrm{e}^{-\lambda/q}$,则

$$u_ip_{ij}=\frac{1}{i!}\left(\frac{\lambda}{q}\right)^i\sum_{k=0}^{\min(i,j)}\mathrm{C}_i^k(1-q)^kq^{i-k}\frac{\lambda^{j-k}}{(j-k)!}\mathrm{e}^{-(1+\lambda/q)}$$

关于 i,j 是对称的,因此 $u_ip_{ij}=u_jp_{ji}$. 所以,链是可逆的,且平稳分布是参数为 λ 的泊松分布.
□

习　题

5-1. 设转移概率矩阵为

(1) $\begin{pmatrix}0&1&0\\1/4&1/2&1/4\\0&1&0\end{pmatrix}$;　　(2) $\begin{pmatrix}1/4&0&1/4&1/2\\1/4&1/4&1/4&1/4\\0&0&1/2&1/2\\1/2&0&1/2&0\end{pmatrix}$,

求平稳分布.

5－2. 设袋中有 a 个球,球为黑色的或白色的.每次从袋中任取一球,然后放回一个不同颜色的球.以 X_n 记取了 n 次球之后袋中白球的个数,则 $\{X_n, n\geqslant 0\}$ 为正常返不可约马尔可夫链,求其平稳分布.这个马尔可夫链称为埃伦弗斯特(Ehrenfest)扩散模型.

5－3. a 个黑球与 a 个白球分装在两个袋子中,每个袋中各 a 个.每次从每个袋中各任取一球,互相交换后放回袋中.以 X_n 记交换 n 次之后第一个袋中黑球的个数,则 $\{X_n, n\geqslant 0\}$ 为遍历不可约马尔可夫链,求其平稳分布.

5－4. 设转移概率为

$$p_{0i}=a_i>0,\quad i\geqslant 0;\quad p_{i,i-1}=1,\quad i\geqslant 1,$$

证明:(1) 链是常返不可约的,且不变测度为 $\{c\sum_{n=i}^{\infty}a_n, i\geqslant 0\}$,其中 c 是正常数;(2) 零常返的充要条件是 $\mu=\sum_{n=1}^{\infty}na_{n-1}=\infty$;(3) 正常返的充要条件是 $\mu=\sum_{n=1}^{\infty}na_{n-1}<\infty$,且这时平稳分布为 $\{\pi_i=\frac{1}{\mu}\sum_{n=1}^{\infty}a_n, i\geqslant 0\}$.

5－5. 设 $X=\{X_n, n\geqslant 0\}$ 为不可约马尔可夫链,令

$$Y_n=\{X_n, X_{n+1}\},\qquad n\geqslant 0,$$

则 $Y=\{Y_n, n\geqslant 0\}$ 是以 $E_Y=\{(i,j): p_{ij}>0\}$ 为状态空间的不可约马尔可夫链.若 X 又是常返(正常返或零常返)的,则 Y 相应地也是常返(正常返或零常返)的.

5－6. 设马尔可夫链的转移概率为

$$p_{ij}=\lambda_i\delta_{ij}+(1-\lambda_i)u_j,\qquad i,j\in E,$$

$$u_j\geqslant 0,\ \sum_j u_j=1,\quad 0<\lambda_i<1,\quad i,j\in E.$$

证明:(1) $A=\{i: u_i>0\}$ 为一常返类;

(2) $E\setminus A$ 中状态全为非常返状态,且从非常返状态出发,迟早必定要进入 A;

(3) A 为正常返类的充要条件是 $\mu=\sum_i\frac{u_i}{1-\lambda_i}<\infty$,且这时平稳分布为

$$\pi_i=\frac{1}{\mu}\frac{u_i}{1-\lambda_i},\qquad i\in E.$$

5－7. 设马尔可夫链的转移概率矩阵为

$$\mathbf{P}=\begin{pmatrix} r_0 & p_0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ q_2 & 0 & r_2 & p_2 & \cdots & 0 & 0 & 0 & \cdots \\ \vdots & & & & \ddots & & & \vdots & \\ q_n & 0 & 0 & 0 & \cdots & r_n & p_n & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

$$p_i>0, q_{i+1}>0, p_i+r_i+q_i=1,\quad i\geqslant 0(q_0=0).$$

证明:(1) $\sum_{n=1}^{\infty}\frac{q_n}{p_n}<\infty$时,链为非常返不可约的;

(2) $\sum_{n=1}^{\infty}\frac{q_n}{p_n}=\infty$, $\sum_{n=1}^{\infty}Q_n=\infty$时,链为零常返不可约的,其中

$$Q_n=\frac{p_0\cdots p_{n-1}}{(1-r_1)\cdots(1-r_n)},\quad n\geqslant 1;$$

(3) $Q=1+\sum_{n=1}^{\infty}Q_n<\infty$时,链为正常返不可约的,且这时平稳分布为 $\pi_n=\frac{Q_n}{Q}, n\geqslant 0$;

(4) 存在不变测度的充要条件为 $\sum_n(1-p_n)=\infty$,且这时不变测度为 $\mu_i=cQ_i, i\in E$,其中 c 为任意常数.

5-8. 设$\{X_n, n\geqslant 0\}$为遍历不可约马尔可夫链,$\{\xi_n, n\geqslant 1\}$为一列 i.i.d. 随机变量,其共同分布为

$$\mathrm{P}(\xi_n=k)=a_k,\qquad k\geqslant 1,$$

且$\{X_n, n\geqslant 0\}$与$\{\xi_n, n\geqslant 1\}$相互独立.令

$$\eta_0=0,\quad \eta_n=\xi_1+\cdots+\xi_n,\quad n\geqslant 1,$$
$$Y_n=X_{\eta_n},\quad n\geqslant 0,$$

则$\{Y_n, n\geqslant 0\}$为正常返不可约马尔可夫链,其转移概率矩阵为

$$\widetilde{\mathbf{P}}=\sum_{k=1}^{\infty}a_k\mathbf{P}^k,$$

且有与$\{X_n, n\geqslant 0\}$相同的平稳分布.

5-9. 设$\{X_n, n\geqslant 0\}$为正常返不可约马尔可夫链.G 为状态空间E 的一个非空真子集.定义

$$T_0=\inf\{n\geqslant 0: X_n\in G\},$$
$$T_n=\inf\{n>T_{n-1}: X_n\in G\}, n\geqslant 1,$$
$$Y_n=X_{T_n},\qquad n\geqslant 0,$$

则$\{Y_n, n\geqslant 0\}$也为正常返不可约马尔可夫链,其转移概率矩阵为

$$\widetilde{\mathbf{P}}=\mathbf{Q}+\mathbf{U}\Big(\sum_{n=0}^{\infty}\mathbf{R}^n\Big)\mathbf{V},$$

其中

$$\mathbf{Q}=(p_{ij})_{i,j\in G},\qquad \mathbf{U}=(p_{ij})_{i\in G, j\in \bar{G}},$$
$$\mathbf{V}=(p_{ij})_{i\in \bar{G}, j\in G},\qquad \mathbf{R}=(p_{ij})_{i,j\in \bar{G}},$$

$\bar{G}=E\setminus G$,且平稳分布为 $\pi_i/\sum_{j\in G}\pi_j, i\in G$.

5-10. 设链为不可约的,F 为状态空间E 的有限子集.令

$$r(F)=\inf\{n\geqslant 1: X_n\in F\}.$$

证明:若对一切 $i\in F$, $\mathsf{E}_i[r(F)]<\infty$,则链为正常返的.

5-11. 证明:不可约链为非常返的充要条件是存在非负数列 $\{u_i, i\in E\}$,使得

$$u_i \geqslant \sum_j u_j p_{ji}, \qquad i\in E,$$

且上述不等式中至少有一个成立不等号.

5-12. 不可约马尔可夫链是可配称的充要条件为对一切 $i,j\in E$ 及任意的 $i_1,\cdots,i_n\in E$,

(1) $p_{ii_1}p_{i_1i_2}\cdots p_{i_nj}>0 \Rightarrow p_{ji_n}p_{i_ni_{n-1}}\cdots p_{i_1i}>0$;

(2) $r_{ij}=\dfrac{p_{ji_n}p_{i_ni_{n-1}}\cdots p_{i_1i}}{p_{ii_1}p_{i_1i_2}\cdots p_{i_nj}}$ 只依赖于 i,j.

§2.6 转移概率的极限性质

在本节中我们要讨论转移概率的极限性质,进一步揭示平稳分布的概率意义.

为今后使用方便起见,我们将下述有关在非负级数中取极限的结果写成一条引理.

引理 6.1 设对一切 $i\in E$ 及 $n\geqslant 1$, $a_i^{(n)}\geqslant 0$,则

$$\varliminf_{n\to\infty}\sum_i a_i^{(n)} \geqslant \sum_i \varliminf_{n\to\infty} a_i^{(n)}. \tag{1}$$

证 任取正整数 N,在下式中令 $n\to\infty$:

$$\sum_i a_i^{(n)} \geqslant \sum_{i=0}^{N} a_i^{(n)},$$

$$\varliminf_{n\to\infty}\sum_i a_i^{(n)} \geqslant \sum_{i=0}^{N} \varliminf_{n\to\infty} a_i^{(n)}.$$

再令 $N\to\infty$ 即得(1)式. □

定理 6.1 设 i 为常返状态,则

$$\lim_{n\to\infty} p_{ii}^{(nd_i)} = d_i\pi_i. \tag{2}$$

证 按照处理周期状态的简化方法,把 $\mathbf{P}^{d_i}$ 看作一个新的马尔可夫链的转移概率矩阵,则在新的链中,i 仍为常返状态,且平均返回时间为 $m_{ii}/d_i = 1/(d_i\pi_i)$(原来的 d_i 步只相当于一步),所以不妨设 $d_i=1$. 我们也不妨假设链是不可约的,这相当于局限在一个常返类中讨论.

先设 $X=\{X_n, n\geqslant 0\}$ 为正常返非周期不可约链,且 $\mathsf{P}(X_0=i)=1$,这里 i 是取定的一个状态. 取马尔可夫链 $Y=\{Y_n, n\geqslant 0\}$ 与 X 独立,但有相同的转移概率矩阵,且 $\mathsf{P}(Y_0=k)=\pi_k$. 令

$$Z_n=(X_n, Y_n), \qquad n\geqslant 0.$$

此时称 $Z=\{Z_n, n\geqslant 0\}$ 为 X 和 Y 的(独立)耦合. 在例 2.4 中,我们已知 $Z=\{Z_n, n\geqslant 0\}$ 是非周期不可约马尔可夫链,其转移概率为

$$p_{(j,k)(l,m)} = p_{jl}p_{km}, \qquad (j,k),(l,m)\in E\times E.$$

因此,不难直接验证 $\{\pi_{(j,k)}=\pi_j\pi_k, (j,k)\in E\times E\}$ 是它的平稳分布. 所以 Z 是正常返的. 令

$$r=\inf\{n\geqslant 0: Z_n=(i,i)\},$$

则 $\mathrm{P}(r<\infty)=1$. 对任意的 j,

$$\begin{aligned}
\mathrm{P}(X_n=j,r\leqslant n)&=\sum_{m=0}^{n}\mathrm{P}(X_n=j,r=m)=\sum_{k}\sum_{m=0}^{n}\mathrm{P}(Z_n=(j,k),r=m)\\
&=\sum_{k}\sum_{m=0}^{n}\mathrm{P}(r=m)\mathrm{P}(Z_n=(j,k)\mid Z_m=(i,i))\\
&=\sum_{k}\sum_{m=0}^{n}\mathrm{P}(r=m)p_{ij}^{(n-m)}p_{ik}^{(n-m)}=\sum_{m=0}^{n}\mathrm{P}(r=m)p_{ij}^{(n-m)}.
\end{aligned}$$

同理有

$$\begin{aligned}
\mathrm{P}(Y_n=j,r\leqslant n)&=\sum_{k}\sum_{m=0}^{n}\mathrm{P}(Z_n=(k,j),r=m)\\
&=\sum_{k}\sum_{m=0}^{n}\mathrm{P}(r=m)\mathrm{P}(Z_n=(k,j)\mid Z_m=(i,i))\\
&=\sum_{m=0}^{n}\mathrm{P}(r=m)p_{ij}^{(n-m)}.
\end{aligned}$$

因此 $\mathrm{P}(X_n=j,r\leqslant n)=\mathrm{P}(Y_n=j,r\leqslant n)$.

$$\begin{aligned}
|p_{ij}^{(n)}-\pi_j|&=|\mathrm{P}(X_n=j)-\mathrm{P}(Y_n=j)|\\
&=|\mathrm{P}(X_n=j,r>n)-\mathrm{P}(Y_n=j,r>n)|\leqslant\mathrm{P}(r>n)\to 0,\quad n\to\infty.
\end{aligned}$$

最后的不等式来自于 $0\leqslant\mathrm{P}(X_n=j,r>n),\mathrm{P}(Y_n=j,r>n)\leqslant\mathrm{P}(r>n)$.

现在设 X 为零常返非周期不可约链. 我们要证明

$$\lim_{n\to\infty}p_{ij}^{(n)}=0,\qquad i,j\in E.\tag{3}$$

在上面的证明中构造马尔可夫链 Z 时,对 Z 的初始分布作一点修改,取 $\mathrm{P}(Y_0=k)=1$,这里 k 是任一取定的状态. 这是必需的,因为现在不存在平稳分布了. 但是 Z 仍然是不可约链. 如果 Z 是非常返的,由于

$$\lim_{n\to\infty}p_{(j,j)(l,l)}^{(n)}=\lim_{n\to\infty}[p_{jl}^{(n)}]^2=0,$$

即得(3)式. 所以,下面我们假设 Z 是常返的(这与初始分布无关). 现在用前面的证明仍可证得:对一切 $i,j,k\in E$,

$$\lim_{n\to\infty}|p_{ij}^{(n)}-p_{kj}^{(n)}|=0.\tag{4}$$

用反证法. 设(3)不成立. 由于 $0\leqslant p_{ij}^{(n)}\leqslant 1$,用对角线法则可知,存在一个子列 $\{n_\upsilon\}$,使得对一切 $j,l\in E$,存在极限

$$\lim_{\upsilon\to\infty}p_{jl}^{(n_\upsilon)}=v_l,\tag{5}$$

且 $\{v_l,l\in E\}$ 不恒为零. 而(5)中的极限与 j 无关则由(4)而来. 由引理 6.1,

$$\sum_{l}v_l\leqslant\lim_{\upsilon\to\infty}\sum_{l}p_{jl}^{(n_\upsilon)}=1.\tag{6}$$

同样由于 $\sum_l p_{jl}=1$，在

$$p_{jm}^{(n_\upsilon+1)}=\sum_l p_{jl}p_{lm}^{(n_\upsilon)}$$

中令 $\upsilon\to\infty$ 可得

$$\lim_{\upsilon\to\infty}p_{jm}^{(n_\upsilon+1)}=\upsilon_m.$$

再在

$$p_{jm}^{(n_\upsilon+1)}=\sum_l p_{jl}^{(n_\upsilon)}p_{lm}$$

中令 $\upsilon\to\infty$，由引理 6.1

$$\upsilon_m\geqslant\sum_l \upsilon_l p_{lm},\qquad m\in E. \tag{7}$$

然而又由于

$$1\geqslant\sum_m \upsilon_m\geqslant\sum_m\sum_l \upsilon_l p_{lm}=\sum_l \upsilon_l\sum_m p_{lm}=\sum_l \upsilon_l,$$

因此(7)式中必须全部成立等号，即$\{\upsilon_j,j\in E\}$是不变测度. 但由(6)知存在平稳分布. 这显然与零常返的假设矛盾. □

下面的定理全面地总结了 $p_{ij}^{(n)}$ 的极限性质.

定理 6.2 (1) 设 j 为非常返状态，则对一切 $i\in E$，

$$\lim_{n\to\infty}p_{ij}^{(n)}=0.$$

(2) 设 j 为常返状态，

(i) 若 $i\notin C(j)$，且 i 是本质状态，则对一切 n，$p_{ij}^{(n)}=0$；

(ii) 若 $i\in C(j)$，且 $j\in C_r(i)$，则

$$\lim_{n\to\infty}p_{ij}^{(nd_j+r)}=\frac{d_j}{m_{jj}},$$

而 $k\neq nd_j+r$（n 为整数）时，$p_{ij}^{(k)}=0$；

(iii) 若 i 为非本质状态，则对 $1\leqslant r\leqslant d_j$，

$$\lim_{n\to\infty}p_{ij}^{(nd_j+r)}=\frac{d_j}{m_{jj}}\sum_{\upsilon=0}^{\infty}f_{ij}^{(\upsilon d_j+r)}. \tag{8}$$

证 (1)及(2)(i)都是显然的. 我们只要证明(2)(ii)及(2)(iii)中的两个极限等式. 我们有

$$p_{ij}^{(nd_j+r)}=\sum_{\upsilon=0}^{n}f_{ij}^{(\upsilon d_j+r)}p_{jj}^{(nd_j-\upsilon d_j)},$$

由(2)式及引理 3.1 得(8)式. 若 $i\in C(j)$，且 $j\in C_r(i)$，则 $k\neq \upsilon d_j+r$ 时，$p_{ij}^{(k)}=0$，从而 $f_{ij}^{(k)}=0$，$\sum_{\upsilon=0}^{\infty}f_{ij}^{(\upsilon d_j+r)}=\sum_{n=1}^{\infty}f_{ij}^{(n)}=1$，即得(2)(ii). □

定理 6.2 的叙述如此复杂，主要是由于状态 j 可能为周期状态引起的. 若 j 为非周期状态，

结果就十分简洁了.

推论 设 j 为非周期状态,则对一切 $i\in E$,

$$\lim_{n\to\infty} p_{ij}^{(n)} = f_{ij}\pi_j. \tag{9}$$

证 检查定理 6.2 的各条,它们都归结为(9). □

若链为遍历不可约的,由定理 6.2 的推论,对一切 $i,j\in E$,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j,$$

即如果转移的步数很大,系统进入状态 j 的概率与初始状态无关.更进一步,不论初始分布如何,我们总有

$$\lim_{n\to\infty}\mathsf{P}(X_n=j)=\lim_{n\to\infty}\sum_i \mathsf{P}(X_0=i)p_{ij}^{(n)}=\sum_i \mathsf{P}(X_0=i)\pi_j=\pi_j,$$

即长时间之后,系统的状态的分布趋于稳定,收敛于平稳分布,平稳分布就是马尔可夫链的极限分布.在实际应用中,在系统运行长时间之后,平稳分布就可以近似地看作为系统状态的分布.这是平稳分布的又一重要概率意义.下面的定理更深入地说明了平稳分布的极限意义.

定理 6.3 设链为正常返不可约的,则对任意的状态 i

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}1_{\{X_k=i\}}=\pi_i\right)=1. \tag{10}$$

证 如定理 3.9,定义 $r_i(0)=0$,

$$r_i(k)=\inf\{n>r_i(k-1):X_n=i\},\quad k\geqslant 1,$$
$$w_i(k)=r_i(k)-r_i(k-1),\quad k\geqslant 1.$$

$\{w_i(k),k\geqslant 2\}$ 是 i.i.d.序列,且 $\mathsf{E}[w_i(2)]=m_{ii}<\infty$,由 i.i.d.序列的强大数定律,

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}r_i(n)=m_{ii}\right)=1. \tag{11}$$

记 $N_n=\sum_{k=1}^{n}1_{\{X_k=i\}}$,则由常返性,$N_n\to\infty$,且

$$r_i(N_n)\leqslant n<r_i(N_n+1),$$
$$\frac{r_i(N_n)}{N_n}\leqslant\frac{n}{N_n}<\frac{r_i(N_n+1)}{N_n}. \tag{12}$$

由(11)及(12)即得(10). □

(10)式中的 $\frac{1}{n}\sum_{k=1}^{n}1_{\{X_k=i\}}$ 就是状态 i 在 $\{X_1,\cdots,X_n\}$ 中出现的频率.这个频率依概率 1 收敛于长时间后系统处于状态 i 的概率 π_i.自然,这也可称之为正常返不可约马尔可夫链的大数定律.

例 6.1 有两个反射壁的随机游动 设状态空间为 $\{0,1,\cdots,a\}$,转移概率矩阵为

$$\mathbf{P}=\begin{pmatrix} r_0 & p_0 & 0 & \cdots & \cdots & \cdots & 0 \\ q_1 & r_1 & p_1 & \cdots & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & q_{a-1} & r_{a-1} & p_{a-1} \\ 0 & \cdots & \cdots & \cdots & 0 & q_a & r_a \end{pmatrix},$$

$$p_i>0, 0\leqslant i\leqslant a-1, \quad q_i>0, 1\leqslant i\leqslant a.$$

这时我们可以设想,在“$-1/2$”及“$a+(1/2)$”两处各有一反射壁,因此这是一个有两个反射壁的对称随机游动.不难看出,这是一个有限不可约链,因而是正常返链.为了求平稳分布,只要解方程组

$$\begin{cases} z_0=r_0z_0+q_1z_1, \\ z_i=p_{i-1}z_{i-1}+r_iz_i+q_{i+1}z_{i+1}, \quad 1\leqslant i\leqslant a-1, \\ z_a=p_{a-1}z_{a-1}+r_az_a. \end{cases}$$

与例 5.4 类似地可得

$$\pi_0=\left(1+\sum_{i=0}^{a-1}\frac{p_i\cdots p_0}{q_{i+1}\cdots q_1}\right)^{-1}, \pi_n=\frac{p_{n-1}\cdots p_0}{q_n\cdots q_1}\pi_0, \quad 1\leqslant n\leqslant a.$$

不难直接验证,这时链是可逆的.

若随机游动是对称的:$p_i=q_{i+1}=1/2, 0\leqslant i\leqslant a-1$,则链是非周期的,因而是遍历的,且平稳分布为均匀分布:$\pi_i=\dfrac{1}{a+1}, 0\leqslant i\leqslant a$.所以,不论游动从哪一个位置出发,长时间之后每个位置都是等可能的:

$$\lim_{n\to\infty}\mathrm{P}(X_n=i)=\frac{1}{a+1}, \quad 0\leqslant i\leqslant a. \qquad \square$$

例 6.2 某人为了雨天上下班的方便,购置了 N 把伞,分别放在家里与办公室里.如出门(不论上班或下班)天晴,他就不带伞.如出门遇天雨,手边也有伞,他就带一把伞用.只有出门遇天雨,但手边又无伞可用,才遇到困境.每次出门,天是否下雨是随机的.出门遇到天雨而又无伞可用,自然也是随机事件,且遇上困境的概率 α 与伞的把数 N 有关.如果要使 $\alpha\leqslant 0.05$,应购多少把伞?

为使问题简化,假设出门遇天雨的概率为 $p(0<p<1)$,且各次出门是否天雨互不相干,自然也与手边有几把伞无关,以 X_n 记出门 n 次之后,此人手边可用的伞数.不难看出,在我们的假设之下,$\{X_n, n\geqslant 0\}$ 是一个有限马尔可夫链,其状态空间为 $\{0,1,\cdots,N\}$,转移概率为

$$p_{0N}=1; \quad p_{j,N-j+1}=p, \quad p_{j,N-j}=q=1-p, \quad 1\leqslant j\leqslant N.$$

事实上,只需注意,若手边无伞,只能空手出门,到另一边之后(不论到办公室或回到家里)手边就有 N 把伞了;若手边有 $j\geqslant 1$ 把伞,出门天雨就带一把伞过去,到另一边之后,手边就有 $N-j+1$ 把伞了,而出门天晴的话,到另一边之后,手边就是 $N-j$ 把伞了.写出转移概率矩阵用起来会更方便.

$$
\mathbf{P}=\begin{pmatrix}0&0&0&\cdots&0&0&1\\0&0&0&\cdots&0&q&p\\0&0&0&\cdots&q&p&0\\\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\q&p&0&\cdots&0&0&0\end{pmatrix}.
$$

易见,这是一个遍历不可约链.

我们首先要知道手边无伞的概率,即 $\mathrm{P}(X_n=0)$. 长时间后,这概率与初始分布无关. 因此,我们只需求它的极限,即 π_0. 为此,列出平稳分布满足的方程组:

$$
\begin{cases}\pi_0=q\pi_N,\\ \pi_k=a\pi_{N-k}+p\pi_{N-k+1}, & 0<k<N,\\ \pi_N=\pi_0+p\pi_1.\end{cases}
$$

由第一个方程, $\pi_N=\dfrac{1}{q}\pi_0$. 再由最后一个方程得

$$
\pi_1=\frac{1}{p}(\pi_N-\pi_0)=\frac{1}{p}\left(\frac{1}{q}-1\right)\pi_0=\frac{1}{q}\pi_0.
$$

接着可求得 $\pi_{N-1}=\dfrac{1}{q}\pi_0$. 一般地,不难验证 $\pi_k=\dfrac{1}{q}\pi_0$, $1\leqslant k\leqslant N$. 因此

$$
1=\pi_0+\cdots+\pi_N=\pi_0+\frac{N}{q}\pi_0,\qquad \pi_0=\frac{q}{q+N}.
$$

按假设,天是否下雨与手边的伞数无关. 所以,出门遇天雨而手边又无伞的概率为 $\alpha=\dfrac{pq}{q+N}$. 一般,天雨的概率 p 未知. 我们可用下列方便的估计($pq\leqslant 1/4$):

$$
\frac{pq}{q+N}\leqslant\frac{1}{4N}.
$$

要使 $\alpha\leqslant 0.05$, 可取 $N=5$, 即购置 5 把伞就够了. 但要使 $\alpha\leqslant 0.01$, 必须取 $N=25$, 这数量就略嫌多了一些. □

对周期状态 j, (9)式虽不再成立,但如果我们讨论平均值 $\frac{1}{n}\sum_{v=1}^{n}p_{ij}^{(v)}$ 代替 $p_{ij}^{(n)}$, 就能消除周期的影响.

引理 6.2 设数列 $\{a_n, n\geqslant 1\}$ 满足条件:存在正整数 d 使

$$
\lim_{k\to\infty}a_{kd+r}=b_r,\qquad 1\leqslant r\leqslant d,
$$

则

$$
\lim_{n\to\infty}\frac{1}{n}\sum_{v=1}^{n}a_v=\frac{1}{d}(b_1+\cdots+b_d).
$$

证 由引理 3.1 可知

$$\lim_{k\to\infty}\frac{1}{k}\sum_{v=1}^{k}a_{vd+r}=b_r,\qquad 1\leqslant r\leqslant d.$$

记 $a=\sup_{n\geqslant 1}|a_n|$,把 n 写成 $n=kd+s,1\leqslant s\leqslant d$,则

$$\begin{aligned}&|\frac{1}{n}\sum_{v=1}^{n}a_v-\frac{1}{d}(b_1+\cdots+b_d)|\\&\leqslant|\frac{1}{n}\sum_{v=kd+1}^{kd+s}a_v|+|\frac{1}{n}\sum_{v=1}^{kd}a_v-\frac{1}{d}(b_1+\cdots+b_d)|\\&\leqslant\frac{da}{n}+\frac{1}{d}\sum_{r=1}^{d}|\frac{kd}{n}\frac{1}{k}\sum_{v=1}^{k}a_{vd+r}-b_r|.\end{aligned}$$

$n\to\infty$时,$k\to\infty,\dfrac{kd}{n}\to 1$,由此即得引理结论. □

定理 6.4 对任意的 $i,j\in E$,存在极限

$$\pi_{ij}=\lim_{n\to\infty}\frac{1}{n}\sum_{v=1}^{n}p_{ij}^{(v)}=f_{ij}\pi_j. \tag{13}$$

证 若 j 为非常返状态,则 $\lim_{n\to\infty}p_{ij}^{(n)}=0$ 及 $\pi_j=0$,(13)式显然成立.

若 j 为常返状态,由(8)式

$$\lim_{k\to\infty}p_{ij}^{(kd_j+r)}=d_j\pi_j\sum_{v=0}^{\infty}f_{ij}^{(vd_j+r)},\quad 1\leqslant r\leqslant d.$$

由引理 6.2

$$\lim_{n\to\inf}\frac{1}{n}\sum_{v=1}^{n}p_{ij}^{(v)}=\frac{1}{d_j}\sum_{r=1}^{d_j}d_j\pi_j\sum_{v=0}^{\infty}f_{ij}^{(vd_j+r)}=f_{ij}\pi_j.\quad □$$

由定理 6.2 及 6.4 可见,只有对正常返状态 i,极限$\overline{\lim}_{n\to\infty}p_{ij}^{(n)}$为正,对其它的状态,这极限都为零.这也可看作正常返与零常返的名称中正与零的由来.

例 6.3 对称随机游动 在例3.1中,我们已经知道,直线与平面上的对称随机游动是常返的.那里也分别计算得到

$$p_{00}^{(n)}\simeq\frac{1}{\sqrt{\pi n}},\qquad p_{(0,0)(0,0)}^{(n)}\simeq\frac{1}{\pi n}.$$

因此

$$\lim_{n\to\infty}p_{00}^{(n)}=0,\qquad \lim_{n\to\infty}p_{(0,0)(0,0)}^{(n)}=0.$$

所以,直线与平面上的对称随机游动都是零常返的. □

例 6.4 设转移概率矩阵为

$$\mathbf{P}=\begin{pmatrix}1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6\\ 0 & 1/4 & 1/4 & 1/2 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 1/3 & 2/3\end{pmatrix},$$

我们来计算矩阵(π_{ij}). 首先讨论状态的分类:$\{1,2,3\}$及$\{4,5\}$是两个本质类,即正常返类,状态0是非本质的.先分别在两个正常返类中求出平稳分布.

$$\begin{cases}\pi_1=\dfrac{1}{4}\pi_1+\pi_2,\\[2mm] \pi_2=\dfrac{1}{4}\pi_1+\pi_3,\\[2mm] \pi_3=\dfrac{1}{2}\pi_1.\end{cases}\qquad \begin{cases}\pi_4=\dfrac{1}{3}\pi_5,\\[2mm] \pi_5=\pi_4+\dfrac{2}{3}\pi_5.\end{cases}$$

由此不难解得

$$\pi_2=\frac{3}{4}\pi_1,\quad \pi_3=\frac{1}{2}\pi_1,\quad \left(1+\frac{3}{4}+\frac{1}{2}\right)\pi_1=1,$$

$$\pi_1=\frac{4}{9},\quad \pi_2=\frac{1}{3},\quad \pi_3=\frac{2}{9}.$$

类似地,

$$\pi_5=3\pi_4,\quad (1+3)\pi_4=1,\quad \pi_4=\frac{1}{4},\quad \pi_5=\frac{3}{4}.$$

现在再计算(f_{ij}). 只要计算$f_{01}=f_{02}=f_{03}$及$f_{04}=f_{05}$. 事实上,尽管$f_{00}=1/6$是易见的,但$\pi_0=0,\pi_{00}=0$,所以不必去求f_{00}的值.其它的f_{ij}或为1或为0,容易按分类确定.

$$f_{01}=\frac{1}{6}f_{01}+\frac{1}{2},\qquad f_{01}=\frac{3}{5},$$

$$f_{04}=\frac{1}{6}f_{04}+\frac{1}{3},\qquad f_{04}=\frac{2}{5}.$$

最后即得

$$(\pi_{ij})=\begin{pmatrix}0 & 4/15 & 1/5 & 2/15 & 1/5 & 3/10\\ 0 & 4/9 & 1/3 & 2/9 & 0 & 0\\ 0 & 4/9 & 1/3 & 2/9 & 0 & 0\\ 0 & 4/9 & 1/3 & 2/9 & 0 & 0\\ 0 & 0 & 0 & 0 & 1/4 & 3/4\\ 0 & 0 & 0 & 0 & 1/4 & 3/4\end{pmatrix}.\quad \square$$

定理 6.5 零常返类必包含无穷多个状态.

证 设零常返类 C 为有限集. 由(10),对任意的 $i,j\in C$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\upsilon=1}^{n}p_{ij}^{(\upsilon)}=0,$$

从而

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\upsilon=1}^{n}\Big(\sum_{j\in C}p_{ij}^{(\upsilon)}\Big)=0.$$

但是 C 是常返类,是闭集,因此 $\sum_{j\in C}p_{ij}^{(\upsilon)}=1$,得出矛盾,故 C 为无限集. □

由定理 6.5 立即可知,有限马尔可夫链没有零常返状态. 这是我们早已知道了的(定理 5.4 的推论).

定理 6.6 存在平稳分布的充要条件是存在正常返类. 若正常返类记为 $D_1,D_2,\cdots$,则平稳分布具有下列形式:

$$u_i=\begin{cases}\lambda_n\pi_i, & i\in D_n,\\ 0, & i\notin\bigcup_n D_n,\end{cases}\tag{14}$$

其中 $\lambda_n\geqslant 0$, $\sum_n\lambda_n=1$.

证 对每个正常返类 D_n,

$$\sum_{j\in D_n}\pi_j=1,\quad \pi_j=\sum_{k\in D_n}\pi_k p_{kj},\qquad j\in D_n.$$

因此,若存在正常返类,可直接验证(14)定义的$\{u_i,i\in E\}$确实为平稳分布.

现在假设$\{u_i,i\in E\}$为平稳分布,要证必存在正常返类,且$\{u_i,i\in E\}$有(14)的形式. 由于

$$u_i=\sum_k u_k\left(\frac{1}{n}\sum_{\upsilon=1}^{n}p_{ki}^{(n)}\right),\qquad i\in E.$$

令 $n\to\infty$,由定理 6.4,

$$u_i=\Big(\sum_k u_k f_{ki}\Big)\pi_i,\qquad i\in E.\tag{15}$$

如果没有正常返状态,则对一切 $i\in E,\pi_i=0,u_i=0$,这与 $\sum_i u_i=1$ 是矛盾的. 所以必存在正常返类. 由(15)式已知,$i\notin\bigcup_j D_j$ 时 $\pi_i=0,u_i=0$. 令 $\lambda_j=\sum_{i\in D_j}u_i$. 设 $i\in D_j$,

$$u_i=\sum_k u_k\left(\frac{1}{n}\sum_{\upsilon=1}^{n}p_{ki}^{(\upsilon)}\right)=\sum_{k\in D_j}u_k\left(\frac{1}{n}\sum_{\upsilon=1}^{n}p_{ki}^{(\upsilon)}\right),$$

在上式中令 $n\to\infty$ 即得 $u_i=\lambda_j\pi_i$. 此外,

$$\sum_j\lambda_j=\sum_j\sum_{i\in D_j}u_i=\sum_j u_i=1.$$

所以,$\{u_i,i\in E\}$具有(14)的形式. □

推论 存在唯一的平稳分布的充要条件是恰有一个正常返类.

在定理 6.5 及 6.6 的证明中我们不难注意到定理 6.4 起到的重要作用. 下面,我们再用它证

明判别常返与正常返的两个准则.

定理 6.7 设不可约链满足下列条件:存在$\{y_i, i\in E\}$及某个状态 j 使得$i\neq j$ 时

$$\sum_k p_{ik}y_k\leqslant y_i$$

及 $\lim_{n\to\infty} y_n = +\infty$,则链为常返的.

证 与定理 4.4 的证明一样,我们构造一个新的马尔可夫链,它使状态 j 为吸收状态,它的转移概率为

$$\tilde{p}_{ik}=\begin{cases}p_{ik}, & i\neq j,\\ \delta_{jk}, & i=j,\end{cases}\qquad i,k\in E.$$

这样对一切 $i\in E$,我们有

$$\sum_k \tilde{p}_{ik}y_k\leqslant y_i. \tag{16}$$

我们可以认为 $y_i\geqslant 0, i\in E$.不然的话,只要用 $y_i-\inf_n y_n$ 代替 y_i,(16)式仍然成立.现在重复使用不等式(16),由归纳法可知,对高阶转移概率 $\tilde{p}_{ij}^{(n)}$也有同样的不等式:

$$\sum_k \tilde{p}_{ik}^{(n)}y_k\leqslant y_i,\qquad i\in E.$$

记 $u_n=\inf_{k\geqslant n}y_k$,则 $\lim_{n\to\infty}u_n=+\infty$,对充分大的 N, $u_N>0$,

$$y_i\geqslant\sum_{k=N+1}^{\infty}\tilde{p}_{ik}^{(n)}y_k\geqslant u_N\sum_{k=N+1}^{\infty}\tilde{p}_{ik}^{(n)},\qquad \frac{y_i}{u_N}\geqslant\sum_{k=N+1}^{\infty}\tilde{p}_{ik}^{(n)},$$

$$\sum_{k=0}^{N}\tilde{p}_{ik}^{(n)}=1-\sum_{k=N+1}^{\infty}\tilde{p}_{ik}^{(n)}\geqslant 1-\frac{y_i}{u_N},$$

$$\sum_{k=0}^{N}\left(\frac{1}{n}\sum_{\nu=1}^{n}\tilde{p}_{ik}^{(\nu)}\right)\geqslant 1-\frac{y_i}{u_N}.$$

令 $n\to\infty$,再令 $N\to\infty$即得

$$\sum_k\tilde{\pi}_{ik}\geqslant 1,\qquad i\in E.$$

然而相反的不等式总是成立的:在式子

$$1=\sum_k\left\{\frac{1}{n}\sum_{\nu=1}^{n}\tilde{p}_{ik}^{(\nu)}\right\}$$

中令 $n\to\infty$,由引理 6.1 即得 $\sum_k\tilde{\pi}_{ik}\leqslant 1$.因此

$$\sum_k\tilde{\pi}_{ik}=1,\qquad i\in E.$$

注意到 $k\neq j$ 时,$\tilde{\pi}_k=0$,因为 k 是非常返的;但 $\tilde{\pi}_j=1$,因为 j 是吸收状态.所以对一切 $i\in E$, $\tilde{f}_{ij}=1$.我们早已知道 $i\neq j$ 时,$f_{ij}=\tilde{f}_{ij}=1$.仍由定理 4.1 的推论可知,原来的不可约链是常返的. □

引理 6.2 设链为正常返不可约的,则对任意 $i\neq j, m_{ij}<\infty$.

证 由于 $j\to i$,存在状态 $i_1,\cdots,i_{n-1}$,使 $p_{ji_1}p_{i_1i_2}\cdots p_{i_{n-1}i}>0$,且 $i_v\neq j, v=1,\cdots,n-1$. 由定理 4.7,

$$m_{jj}=\sum_{k\neq j}p_{jk}m_{kj}+1,$$

由于 $m_{jj}<\infty, p_{ji_1}>0, i_1\neq j$,所以必须 $m_{i_1j}<\infty$. 同样地由于

$$m_{i_1j}=\sum_{k\neq j}p_{i_1k}m_{kj}+1$$

及 $p_{i_1i_2}>0, i_2\neq j$,得到 $m_{i_2j}<\infty$. 依次下去,即可知 $m_{ij}<\infty$. □

定理 6.8 不可约链为正常返的充要条件是存在非负数列 $\{z_i, i\in E\}$ 及一个状态(或任一个状态)j,使得

$$\begin{cases}\sum_k p_{ik}z_k\leqslant z_i-1, & i\neq j,\\ \sum_k p_{jk}z_k<\infty.\end{cases}\tag{17}$$

证 先证必要性. 设链为正常返不可约的. 任取一个状态 j,令

$$z_i=\begin{cases}m_{ij}, & i\neq j,\\ 0, & i=j.\end{cases}$$

仍由定理 4.7, $i\neq j$ 时

$$\sum_k p_{ik}z_k=\sum_{k\neq j}p_{ik}m_{kj}=m_{ij}-1=z_i-1.$$

另一方面,

$$\sum_k p_{jk}z_k=\sum_{k\neq j}p_{jk}m_{kj}=m_{jj}-1<\infty.$$

再证充分性. 这时有一个状态 j 使(17)成立. 记 $\delta=\sum_k p_{jk}z_k<\infty$. 令

$$z_i^{(1)}=z_i, \qquad z_i^{(n+1)}=\sum_k p_{ik}^{(n)}z_k, \quad n\geqslant 1, \quad i\in E.$$

易见, $z_j^{(2)}=\delta, z_i^{(2)}\leqslant z_i-1, i\neq j$. 而 $n\geqslant 1$ 时

$$\begin{aligned}z_i^{(n+2)}&=\sum_k\Big(\sum_l p_{il}^{(n)}p_{lk}\Big)z_k=\sum_l p_{il}^{(n)}z_l^{(2)}\\&\leqslant\delta p_{ij}^{(n)}+\sum_{l\neq j}p_{il}^{(n)}(z_l-1)\\&=(1+\delta)p_{ij}^{(n)}-1+\sum_{l\neq j}p_{il}^{(n)}z_l\\&\leqslant(1+\delta)p_{ij}^{(n)}-1+z_i^{(n+1)}, \quad i\in E.\end{aligned}\tag{18}$$

由(18)式,用归纳法可知,对 $n\geqslant 1$ 及 $i\in E, z_i^{(n)}<\infty$,且

$$0\leqslant z_i^{(n+2)}\leqslant(1+\delta)\sum_{v=1}^{n}p_{ij}^{(v)}-n+z_i^{(2)},$$

$$\frac{1}{n}\sum_{\nu=1}^{n} p_{ij}^{(\nu)} \geqslant \frac{1}{1+\delta}\left(1-\frac{z_i^{(2)}}{n}\right).$$

令 $n\to\infty$,即得 $\pi_{ij}\geqslant\dfrac{1}{1+\delta}>0$,从而 $\pi_j>0$,j 为正常返,链也为正常返的. □

我们对离散时间马尔可夫链的一般讨论,到此告一段落.总的说来,我们对一个马尔可夫链所作的分析,包含这样一些要点:先确定转移概率矩阵,这是讨论的出发点;然后将状态进行分类,确定状态的周期;判别状态类是本质或非本质的,常返或非常返的,正常返或零常返的;求平稳分布或不变测度,注意链是否可逆或可配称.掌握了这些情况,就能对一个马尔可夫链的变化趋势及概率规律有一个充分的认识.对于一个具体的马尔可夫链,往往是某一些问题有突出的意义,应依据具体情况作具体的分析,不一定要面面俱到.对于有限马尔可夫链,情况要简单得多,因为它的状态或为非本质的,或为正常返的,而且必定存在正常返状态,也就必有平稳分布.另一方面,从非本质状态出发,必定迟早要进入正常返状态.

尽管我们对离散时间的马尔可夫链作了相当深入的讨论,但毕竟还只是初步的.特别,由于我们局限于只使用初等的方法,以致无法对马尔可夫链的许多较深刻的概率性质作进一步的讨论.继续讨论马尔可夫链的理论需要进入专门的研究领域.

在应用领域中,在建立数学模型时,常常可能遇到马尔可夫链.在充分讨论了所遇到的马尔可夫链的性质之后,就可以按照实际的需要,解决所提出的问题.下面我们再举两个排队论的例子.对这些例子的分析也是综合运用已讲述过的理论的练习.

例 6.5　排队系统 $M/G/1$ 的嵌入链　设在某个服务部门,顾客随机地相继来到.假定顾客来到的间隔时间构成独立同分布的随机变量.有若干个服务员按照一定的规则依次为顾客服务,通常按照"先来先服务"的规则办事.一个顾客到来时,如有服务员空着,就可以立刻得到服务,服务完毕,顾客随即离去.如果全体服务员都在工作,顾客就按来到的次序排起队来.当有一个服务员结束服务时有顾客在排队,他就立即为排在最前面的顾客服务.假定各个顾客的服务时间也是独立同分布的随机变量,且服务时间与顾客来到间隔相互独立.这样的排队系统记为 $G/G/S$,其中第一个 G 代表相继来到的两个顾客之间的时间间隔的分布,第二个 G 代表一个顾客的服务时间的分布,S 表示有 S 个服务员.实际生活中发生的许多排队现象,都可以归结为上述模型,只是除了通常意义的顾客与服务员外,顾客与服务员还可以有不同的含义.例如飞机到机场降落,这时顾客是飞机,服务员是跑道;汽车到加油站加油,这时顾客是汽车,服务员是加油机等等.

现在假设顾客来到的时间间隔服从指数分布,参数为 λ,即在 $(0,t]$ 中来到的顾客总数是一个参数为 λ 的泊松过程.这时在 $(t,t+\Delta t]$ 中新来到一个顾客的概率为 $\lambda\Delta t+o(\Delta t)$,来到多于一个顾客的概率为 $o(\Delta t)$,$(t,t+\Delta t]$ 中来到的顾客数与 $(0,t]$ 中顾客来到的情况相互独立.设只有 1 个服务员,按先来先服务的原则接待顾客.这样的排队系统记为 $M/G/1$.M 是指数分布的代号.

我们以 X_n 记第 n 个顾客服务完毕离开时系统中的(在接受服务及在排队等候的)顾客总数(也称为队伍长度).初始时刻取为第一个顾客来到的时刻,即 $X_0=1$,则 $\{X_n,n\geqslant 0\}$ 为马尔可夫链.事实上,以 ξ_n 记第 n 个顾客数在接受服务期间来到的顾客数,则

$$X_n=\max(X_{n-1}-1,0)+\xi_n,\qquad n\geqslant 1,$$

且$\{\xi_n, n\geqslant 1\}$为 i.i.d. 随机变量,其共同分布为

$$\mathrm{P}(\xi_n = k) = a_k = \int_0^{\infty} \frac{(\lambda t)^k}{k!} \mathrm{e}^{-\lambda t} \mathrm{d}G(t), \qquad k = 0,1,2,\cdots.$$

这里直观地利用了泊松过程在任一长度为 t 的区间上的增量服从参数为 λt 的泊松分布,且不相交区间上的增量相互独立.由于$\{X_n, n\geqslant 0\}$是一些特定时刻的队长,所以通常被称为嵌入链.由定理 1.5 可知,$\{X_n, n\geqslant 0\}$的转移概率矩阵为

$$\mathbf{P} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

即对一切 $j \in E$,

$$p_{0j} = a_j, \qquad p_{ij} = \begin{cases} a_{j-i+1}, & j \geqslant i-1, \\ 0, & j < i-1, \end{cases} \quad i \geqslant 1.$$

我们有

$$a_0 > 0, \qquad a_0 + a_1 < 1. \tag{19}$$

容易看出,在条件(19)之下,链是非周期不可约的.记 $\rho = \sum_{k=1}^{\infty} k a_k$.我们将证明:

(1) 若 $\rho > 1$,链为非常返的;

(2) 若 $\rho = 1$,链为零常返的;

(3) 若 $\rho < 1$,链为正常返的.

事实上,ρ 是一个顾客的服务期内,来到的顾客的平均数.从直观上不难理解.当 $\rho > 1$ 时,队伍会越来越长,链不会是常返的.我们也早已熟知,链为非常返不可约时,必有

$$\mathrm{P}(\lim_{n\to\infty} X_n = +\infty) = 1.$$

当 $\rho < 1$ 时,队伍自然不会越来越长,队伍的长度近似地有一个稳定的分布,这也是能够从直观上想到的.只有 $\rho = 1$ 的情形难以从直观上作出判断.

设 $\rho > 1$.记 $A(z) = \sum_{k=0}^{\infty} a_k z^k$.由引理 4.4,这时方程 $A(z) = z$ 在$(0,1)$中有一个根 ξ.令 $z_i = \xi^i, i \geqslant 0$,则 $i \geqslant 1$ 时

$$\sum_k p_{ik} z_k = \sum_{k=i-1}^{\infty} a_{k-i+1} \xi^k = \xi^{i-1} \sum_{k=i-1}^{\infty} a_{k-i+1} \xi^{k-i+1}$$
$$= \xi^{i-1} A(\xi) = \xi^i = z_i.$$

$\{z_i = \xi^i, i \geqslant 0\}$显然是非常数的有界数列,故由定理 4.5,链是非常返的.

设 $\rho \leqslant 1$.当 $i \geqslant 1$ 时,

$$
\begin{aligned}
\rho &= \sum_{k=1}^{\infty} k a_k = \sum_{k=i-1}^{\infty} (k-i+1) a_{k-i+1} \\
&= \sum_k k p_{ik} - (i-1) \sum_{k=i-1}^{\infty} a_{k-i+1} \geqslant \sum_k k p_{ik} - (i-1), \qquad (20) \\
i &\geqslant \sum_k k p_{ik} + 1 - \rho \geqslant \sum_k p_{ik} k,
\end{aligned}
$$

令 $y_i = i, i \geqslant 0$，则

$$
y_i \geqslant \sum_k p_{ik} y_k, \qquad i \neq 0,
$$

且 $\lim_{n \to \infty} y_n = \infty$，故由定理 6.6，链是常返的.

设 $\rho < 1$，当 $i \geqslant 1$ 时，由(20)式

$$
\sum_k p_{ik} \frac{k}{1-\rho} \leqslant \frac{i}{1-\rho} - 1.
$$

令 $z_i = \frac{i}{1-\rho}$，则

$$
\sum_k p_{ik} z_k \leqslant z_i - 1, \qquad i \neq 0,
$$

$$
\sum_k p_{0k} z_k = \sum_k a_k \frac{k}{1-\rho} = \frac{\rho}{1-\rho} < \infty,
$$

故由定理 6.7，链为正常返的.

最后，假设链为正常返的，我们要证明 $\rho < 1$，且同时求得平稳分布的母函数 $\pi(z) = \sum_{k=0}^{\infty} \pi_k z^k$. 将式子

$$
\pi_i = \sum_k \pi_k p_{ki} = \pi_0 a_i + \sum_{k=1}^{i+1} \pi_k a_{i-k+1}
$$

两端乘以 z^i，再对 i 相加可得

$$
\begin{aligned}
\pi(z) &= \pi_0 A(z) + \sum_{i=0}^{\infty} \Big(\sum_{k=1}^{i+1} \pi_k a_{(i+1)-k} \Big) z^i \\
&= \pi_0 A(z) + \frac{1}{z} \sum_{i=0}^{\infty} \Big(\sum_{k=0}^{i+1} \pi_k a_{(i+1)-k} - \pi_0 a_{i+1} \Big) z^{i+1} \\
&= \pi_0 A(z) + \frac{1}{z} \Big[\sum_{i=1}^{\infty} \Big(\sum_{k=0}^{i} \pi_k a_{i-k} \Big) z^i - \pi_0 \sum_{i=1}^{\infty} a_i z^i \Big] \\
&= \pi_0 A(z) + \frac{1}{z} \Big[\sum_{i=0}^{\infty} \Big(\sum_{k=0}^{i} \pi_k a_{i-k} \Big) z^i - \pi_0 \sum_{i=0}^{\infty} a_i z^i \Big] \\
&= \pi_0 A(z) + \frac{1}{z} [\pi(z) A(z) - \pi_0 A(z)], \qquad (21)
\end{aligned}
$$

从而有

$$
\frac{A(z)-1}{z-1} = 1 - \pi_0 \frac{A(z)}{\pi(z)}.
$$

令 $z \to 1$ 即得

$$\rho = A'(1) = 1 - \pi_0 < 1.$$

另一方面,由(21)可解得

$$\pi(z) = \pi_0 \frac{(1-z)A(z)}{A(z)-z} = \frac{(1-\rho)(1-z)A(z)}{A(z)-z}. \quad \square$$

例 6.6 排队系统 $G/M/1$ 的嵌入链 现在考虑排队系统 $G/M/1$,这时服务时间服从指数分布,参数为 μ.现在以 X_n 记第 $n+1$ 个顾客到达时的队伍长度($X_0 = 1$),则 $\{X_n, n \geqslant 0\}$ 也为马尔可夫链,它是 $G/M/1$ 的嵌入链.事实上,令 ξ_n 为第 n 个顾客与第 $n+1$ 个顾客到达时刻之间的间隔内服务完毕的顾客数,则

$$X_n = \max(X_{n-1} + 1 - \xi_n, 0), \qquad n \geqslant 1,$$

且 $\{\xi_n, n \geqslant 1\}$ 为 i.i.d. 随机变量,其共同分布为

$$\mathsf{P}(\xi_n = k) = a_k = \int_0^\infty \frac{(\mu t)^k}{k!} \mathrm{e}^{-\mu t} \mathrm{d}G(t), \qquad k = 0, 1, 2, \cdots.$$

同样地,由定理 1.5 可知,$\{X_n, n \geqslant 0\}$ 的转移概率矩阵为

$$\mathbf{P} = \begin{pmatrix} \alpha_0 & a_0 & 0 & 0 & \cdots \\ \alpha_1 & a_1 & a_0 & 0 & \cdots \\ \alpha_2 & a_2 & a_1 & a_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

其中 $\alpha_i = \sum_{k=i+1}^\infty a_k$,即

$$p_{ij} = \begin{cases} \alpha_i, & j = 0, \\ a_{i-j+1}, & 0 < j \leqslant i+1, \\ 0, & i+1 < j. \end{cases}$$

条件(19)仍然成立,这时链仍为非周期不可约的.

同样记 $\rho = \sum_{k=1}^\infty k a_k$.我们将证明:

(1) 若 $\rho < 1$,链为非常返的;

(2) 若 $\rho = 1$,链为零常返的;

(3) 若 $\rho > 1$,链为正常返的.

现在 ρ 是两个相继的顾客到达时刻之间的间隔内服务完毕的平均顾客数.类似地,也可给上述结果作直观的解释.

设 $\rho > 1$.由引理 4.4,方程 $A(z) = \sum_{k=0}^\infty a_k z^k = z$ 在(0,1)中有一个根 ξ.令 $z_i = \xi^i, i \geqslant 0$,则 $i \geqslant 1$ 时

$$\sum_k z_k p_{ki} = \sum_{k=i-1}^\infty \xi^k a_{k-i+1} = \xi^{i-1} \sum_{k=0}^\infty a_k \xi^k = \xi^{i-1} A(\xi) = \xi^i = z_i,$$

$$\sum_{k=0}^{\infty} z_k p_{k0} = \sum_{k=0}^{\infty} \xi^k \alpha_k = \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} a_j \xi^k = \sum_{j=1}^{\infty} \sum_{k=0}^{j-1} a_j \xi^k = \sum_{j=1}^{\infty} a_j \frac{1-\xi_j}{1-\xi}$$

$$= \frac{1}{1-\xi} \sum_{j=0}^{\infty} a_j (1-\xi^j) = \frac{1-A(\xi)}{1-\xi} = 1 = z_0.$$

此外，$\sum_{k=0}^{\infty} z_k = \frac{1}{1-\xi} < \infty$，所以链为正常返，且平稳分布为

$$\pi_i = (1-\xi)\xi^i, \qquad i \geqslant 0.$$

设 $\rho \leqslant 1$. $i \geqslant 1$ 时

$$\sum_{k=0}^{\infty} p_{ki} = \sum_{k=i-1}^{\infty} a_{k-i+1} = \sum_{k=0}^{\infty} a_k = 1,$$

$$\sum_{k=0}^{\infty} p_{k0} = \sum_{k=0}^{\infty} \alpha_k = \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} a_j = \sum_{j=1}^{\infty} j a_j = \rho \leqslant 1.$$

令 $z_i = 1, i \geqslant 0$，则

$$z_i \geqslant \sum_k z_k p_{ki}, \qquad i \geqslant 0.$$

但 $\sum_i z_i = \infty$，由定理 5.6，链不是正常返的.

为了进一步区分(零)常返与非常返，讨论线性方程组

$$y_i = \sum_k p_{ik} y_k, \qquad i \neq 0 \tag{22}$$

的解.不妨设 $y_0 = 0$.不然的话，可用 $y_i - y_0$ 代替 y_i. 记 $Y(z) = \sum_{k=1}^{\infty} y_k z^k$. 这时

$$y_i = \sum_{k=1}^{i+1} a_{i+1-k} y_k, \qquad i \neq 0,$$

$$zY(z) = \sum_{i=1}^{\infty} y_i z^{i+1} = \sum_{i=1}^{\infty} \sum_{k=1}^{i+1} a_{i+1-k} y_k z^{i+1} = \sum_{i=2}^{\infty} \Big(\sum_{k=1}^{i} a_{i-k} y_k\Big) z^i$$

$$= \sum_{i=1}^{\infty} \Big(\sum_{k=1}^{i} a_{i-k} y_k\Big) z^i - a_0 y_1 z = A(z) Y(z) - a_0 y_1 z,$$

$$Y(z) = \frac{a_0 y_1 z}{A(z) - z}. \tag{23}$$

另一方面，我们有

$$\frac{A(z)-z}{1-z} = 1 - \frac{1-A(z)}{1-z} = 1 - \Big(1 - \sum_{k=0}^{\infty} a_k z^k\Big)\Big(\sum_{k=0}^{\infty} z^k\Big)$$

$$= 1 - \sum_{k=0}^{\infty} \Big(1 - \sum_{n=0}^{k} a_n\Big) z^k = 1 - \sum_{k=0}^{\infty} \Big(\sum_{n=k+1}^{\infty} a_n\Big) z^k = 1 - W(z),$$

这里

$$W(z) = \sum_{k=0}^{\infty} \Big(\sum_{n=k+1}^{\infty} a_n\Big) z^k, \qquad W(1) = \sum_{k=0}^{\infty} \sum_{n=k+1}^{\infty} a_n = \sum_{n=1}^{\infty} n a_n \leqslant 1,$$

因此

$$\frac{1-z}{A(z)-z}=\frac{1}{1-W(z)}=\sum_{n=0}^{\infty}[W(z)]^n=\sum_{n=0}^{\infty}b_nz^n, \tag{24}$$

展开式中的系数 b_n 非负：$b_n\geqslant 0, n\geqslant 0$.

再由(23)式，

$$Y(z)=a_0y_1z\frac{1-z}{A(z)-z}\frac{1}{1-z}=a_0y_1z\sum_{n=0}^{\infty}b_nz^n\sum_{n=0}^{\infty}z_n$$

$$=a_0y_1\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n}b_k\right)z^{n+1},$$

$$y_{n+1}=a_0y_1\sum_{k=0}^{n}b_k,\quad n\geqslant 0, \tag{25}$$

然而由(24)式

$$\sum_{k=0}^{\infty}b_k=\lim_{z\to 1^-}\frac{1-z}{A(z)-z}=\lim_{z\to 1^-}\frac{-1}{A'(z)-1}$$

$$=\lim_{z\to 1^-}\frac{1}{1-\sum_{k=1}^{\infty}ka_kz^{k-1}}$$

$$=\begin{cases}\dfrac{1}{1-\rho}, & \rho<1,\\ +\infty, & \rho=1.\end{cases}$$

因此，若 $\rho=1$，$\{y_n,n\geqslant 0\}$ 为方程组(22)的有界解，必须 $y_n=0, n\geqslant 1$. 注意到我们假设 $y_0=0$，即用 y_i-y_0 代替 y_i，所以方程组(22)的有界解必为常数，从而链为常返的. 若 $\rho<1$，按(25)式定义 $y_n, n\geqslant 1$，取 $y_1\neq 0$ 及 $y_0=0$，把前面的推导倒推回去，可知 $\{y_n,n\geqslant 0\}$ 为方程组(22)的非常数的有界解，从而链为非常返的. □

习　题

6-1. 设存在 $k\geqslant 1$ 及正常数列 $\{C_j, j\in E\}$，使得对一切 $i,j\in E$，

$$p_{ij}^{(k)}\geqslant C_j>0,$$

证明：链为遍历不可约的.

6-2. 证明：有限马尔可夫链为遍历不可约的充要条件是存在 $k\geqslant 1$，使得对一切 $i,j\in E$，$p_{ij}^{(k)}>0$.

6-3. 若链为有限遍历不可约的，则存在常数 $c>0$ 及 $0<\rho<1$，使得对一切 $i,j\in E$ 及 $n\geqslant 1$，

$$|p_{ij}^{(n)}-\pi_j|\leqslant c\rho^n.$$

6-4. 设市场上有 a 种牌号的牙膏，记为 $\{1,2,\cdots,a\}$. 假设消费者相继选用的牙膏牌号构成马尔可夫链. 选用第 i 种牌号牙膏的消费者继续使用第 i 种牌号牙膏的概率为 p_{ii}，$0<p_{ii}<1$，

$i=1,\cdots,a$. 如果他对原来用的牙膏不满意，就在其它的 $a-1$ 种牙膏中任选一种. 因此

$$p_{ij}=\frac{1-p_{ii}}{a-1},\qquad i\neq j.$$

求长时间后第 i 种牌号的牙膏在市场上的占有率.

6-5. 一个工厂使用 S 部同样的机器. 一旦一部机器出毛病，就予以更新. 在每周开始时，工厂提出新机器的定单，以保持机器的总数为 S，但需等待一周方能收到定货. 以 X_{n-1}表示第 n 周开始时，正常工作的机器数，设 $X_0=S$. 以 Y_n 记第 n 周内出毛病的机器数. 假设

$$\mathrm{P}(Y_n=j\mid X_{n-1})=\frac{1}{n+1},\quad j=0,1,\cdots,i,$$

则$\{X_n,n\geqslant 0\}$为马尔可夫链. 求长时间后，每周开始时正常工作着的机器的平均数.

6-6. 设马尔可夫链有 $a+1$ 个状态$\{0,1,\cdots,a\}$，其中 0 及 a 为吸收状态，$\{1,\cdots,a-1\}$为非常返类. 取定状态 k，$1\leqslant k\leqslant a-1$，$u_k$ 为从 k 出发被 0 或 a 吸收的平均吸收时间. 现在构造一个新的链，其转移概率为

$$\tilde{p}_{0k}=\tilde{p}_{ak}=1;\qquad \tilde{p}_{ij}=p_{ij},\qquad i\neq 0 \text{ 或 } a,$$

则新的链有唯一的平稳分布：$\{\tilde{\pi}_i,0\leqslant i\leqslant a\}$. 证明：

$$u_k=(\tilde{\pi}_0+\tilde{\pi}_a)^{-1}-1.$$

6-7. 设转移概率矩阵为

$$(1)\ \begin{pmatrix} 0 & 1/4 & 1/2 & 0 & 1/4 & 0\\ 1/4 & 0 & 0 & 1/4 & 0 & 1/2\\ 0 & 0 & 1/2 & 1/2 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1/3 & 2/3\\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix};$$

$$(2)\ \begin{pmatrix} 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4\\ 0 & 1/4 & 0 & 1/2 & 0 & 1/4\\ 0 & 0 & 1/2 & 1/2 & 0 & 0\\ 0 & 0 & 0 & 1/2 & 1/2 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

求矩阵(π_{ij}).

6-8. 记 $\boldsymbol{\Pi}=(\pi_{ij})$. 证明：

(1) $\boldsymbol{\Pi}=\boldsymbol{\Pi}\mathbf{P}$;

(2) $\boldsymbol{\Pi}=\mathbf{P}\boldsymbol{\Pi}$;

(3) $\boldsymbol{\Pi}=\boldsymbol{\Pi}^2$;

(4) $(\mathbf{P}-\boldsymbol{\Pi})^n=\mathbf{P}^n-\boldsymbol{\Pi},n\geqslant 1$.

6－9．若对一切 $i \in E$，$\sum_k \pi_{ik}=1$，链称为非消散的．证明：

(1) 链为非消散的充要条件是：本质状态必为正常返状态，且从非本质状态出发迟早要进入正常返状态；

(2) 若 $\inf_i \sum_j \pi_{ij}>0$，则链为非消散的．

6－10．设转移概率矩阵满足条件：

$$\sum_i p_{ij}=1, \qquad j \in E,$$

即正常数列为不变测度(这样的转移概率矩阵称为双随机的)．证明：

(1) 若状态空间有限，则全部状态为正常返的；

(2) 若链为不可约且状态空间有限，则平稳分布为均匀分布；

(3) 若链为不可约且状态空间无穷，则链不是正常返的．

6－11．设马尔可夫链的状态空间为平面上两条坐标轴上的整数点，且转移概率为

$$p_{(0,0)(0,1)}=p_{(0,0)(1,0)}=p_{(0,0)(0,-1)}=p_{(0,0)(-1,0)}=\frac{1}{4},$$

$$p_{(i,0)(i+1,0)}=p, p_{(i,0)(i-1,0)}=1-p, i \neq 0, 0<p<1,$$

$$p_{(0,i)(0,i+1)}=r, p_{(0,i)(0,i-1)}=1-r, i \neq 0, 0<r<1.$$

证明：链或为非常返不可约的，或为零常返不可约的；(零)常返的充要条件是 $p=r=1/2$．

6－12．设转移概率为

$$p_{01}=1, p_{ij}=\frac{i^j}{j!}\mathrm{e}^{-i}, i \geqslant 1.$$

证明：链为零常返不可约的．

6－13．设转移概率为

$$p_{01}=1, \quad p_{i,i-1}=q, \quad p_{i,i+k}=p_k, \quad i \geqslant 1, k \geqslant 0,$$

$$0<q<1, \qquad q+\sum_{k=0}^{\infty} p_k=1.$$

证明：(1) $\sum_{k=1}^{\infty} kp_k>q$ 时，链为非常返不可约的；

(2) $\sum_{k=1}^{\infty} kp_k=q$ 时，链为零常返不可约的；

(3) $\sum_{k=1}^{\infty} kp_k<q$ 时，链为正常返不可约的．

6－14．证明：不可约链是正常返的，若存在非负数列 $\{z_i, i \in E\}$，正数 e 及有限个状态 $j_1, \cdots, j_m$ 使得

$$\begin{cases} \sum_k p_{ik}z_k \leqslant z_i-e, & i \notin \{j_1, \cdots, j_m\}, \\ \sum_k p_{ik}z_k<\infty, & i \in \{j_1, \cdots, j_m\}. \end{cases}$$

第三章　连续时间的马尔可夫链

§3.1　转移概率函数与密度矩阵

在第二章中，我们讨论了离散时间的马尔可夫链，时间离散是指时间参数取非负整数值. 在这一章中，我们要讨论连续时间的马尔可夫链，时间连续是指时间参数取非负实数值.

定义　设 $X=\{X_t,t\geqslant 0\}$ 为一族只取非负整数值的随机变量. 若对任意的 $k\geqslant 1,t_0<t_1<\cdots<t_{k+1}$ 及非负整数 $i_0,i_1,\cdots,i_{k+1}$，

$$\begin{aligned}&\mathrm{P}(X_{t_{k+1}}=i_{k+1}\mid X_{t_0}=i_0,X_{t_1}=i_1,\cdots,X_{t_k}=i_k)\\&\quad=\mathrm{P}(X_{t_{k+1}}=i_{k+1}\mid X_{t_k}=i_k),\end{aligned}\tag{1}$$

则称 $X=\{X_t,t\geqslant 0\}$ 为连续时间离散状态的马尔可夫过程，或连续时间的马尔可夫链.

这里马尔可夫性的定义(1)与离散时间的场合形式上完全一致. 因此在离散时间场合对马尔可夫性所作的讨论(第二章的定理 1.1 与 1.2 以及将来的事件允许涉及可列无穷多个将来时刻等)在连续时间的场合依然有效，我们就不再重复叙述了. 关于状态空间的说明，对连续时间场合也同样适用. 我们仍用 E 记状态空间. 在今后的一般讨论中取 E 为 $\mathbb{N}=\{0,1,2,\cdots\}$.

我们也将只讨论齐次的马尔可夫链，即假设对任意的 $i,j\in E$，转移概率

$$\mathrm{P}(X_{s+t}=j\mid X_s=i)=p_{ij}(t),\quad t\geqslant 0$$

只与 t 有关，与 s 无关. 显然，$p_{ij}(0)=\delta_{ij}$. 易见，科尔莫戈罗夫－查普曼方程也成立：

$$p_{ij}(t+s)=\sum_k p_{ik}(t)p_{kj}(s),\quad t,s\geqslant 0.\tag{2}$$

我们也把转移概率写成矩阵的形式：

$$\mathbf{P}(t)=(p_{ij}(t))_{i,j\in E},\quad t\geqslant 0,$$

从而科尔莫戈罗夫－查普曼方程有下列简单的形式：

$$\mathbf{P}(t+s)=\mathbf{P}(t)\mathbf{P}(s),\qquad t,s\geqslant 0.\tag{3}$$

现在转移概率是 t 的函数，所以 $\mathbf{P}(t)$ 或 $p_{ij}(t)$ 都称为转移概率函数. 我们知道，在离散时间情形，全部高阶的转移概率矩阵由一个(一步)转移概率矩阵完全决定，但在连续时间情形，从科尔莫戈罗夫－查普曼方程推不出类似的结论. $\{\mathbf{P}(t),t\geqslant 0\}$ 不能由一个 $\mathbf{P}(1)$ 或某个 $\mathbf{P}(t_0)(t_0>0)$ 完全决定. 然而设法找到一个矩阵，用它即能决定与刻画 $\{\mathbf{P}(t),t\geqslant 0\}$，将是今后讨论的一个主要课题.

连续时间马尔可夫链的有限维分布同样由初始分布

$$\mathrm{P}(X_0=i)=p_i,\qquad i\in E$$

及转移概率函数完全决定:对任意的 $n\geqslant1,0=t_0<t_1<\cdots<t_n$ 及 $i_0,i_1,\cdots,i_n\in E$,

$$\begin{aligned}&\mathsf{P}(X_{t_0}=i_0,X_{t_1}=i_1,\cdots,X_{t_n}=i_n)\\&=p_{i_0}p_{i_0i_1}(t_1-t_0)\cdots p_{i_{n-1}i_n}(t_n-t_{n-1}).\end{aligned}\tag{4}$$

反过来,若一族随机变量 $\{X_t,t\geqslant0\}$ 的有限维分布由(4)给出,其中 $\{p_i,i\in E\}$ 为一概率分布,$\{\mathbf{P}(t)=(p_{ij}(t))_{i,j\in E},t\geqslant0\}$ 为一族满足条件(3)的随机矩阵,则可直接验证 $\{X_t,t\geqslant0\}$ 为马尔可夫链,$\{p_i,i\in E\}$ 为其初始分布,$\{\mathbf{P}(t),t\geqslant0\}$ 为其转移概率函数.

我们已经知道,对于马尔可夫链,我们主要是研究它的状态转移的规律与性质,即由转移概率所决定的性质.在离散时间情形,我们就把一个随机矩阵当作一个马尔可夫链.在连续时间情形,我们也把一个满足下列三个条件的函数矩阵 $\{\mathbf{P}(t)=(p_{ij}(t))_{i,j\in E},t\geqslant0\}$:

1) $p_{ij}(t)\geqslant0,\quad t\geqslant0,\quad p_{ij}(0)=\delta_{ij},\quad i,j\in E;$ (5)

2) $\sum\limits_j p_{ij}(t)=1,\quad t\geqslant0,\quad i\in E;$ (6)

3) $p_{ij}(t+s)=\sum\limits_k p_{ik}(t)p_{kj}(s),\quad t,s\geqslant0,\quad i,j\in E,$ (7)

看作一个马尔可夫链,$p_{ij}(t)$ 即为它的转移概率函数.满足条件(5)、(6)及(7)的函数族 $\{p_{ij}(t),t\geqslant0,i,j\in E\}$ 也称为转移概率函数.

今后我们需要讨论更广泛的一类函数族,即把条件(6)放宽为

$$\sum_j p_{ij}(t)\leqslant1,\qquad t\geqslant0,\quad i\in E.\tag{8}$$

我们把满足(5)、(8)及(7)的 $\{p_{ij}(t),t\geqslant0,i,j\in E\}$ 称为*广转移概率函数*.这时如果把状态空间 E 扩大,再加进一个新的状态 θ,并定义

$$\begin{aligned}&p_{\theta\theta}(t)=1,\quad p_{\theta i}(t)=0, && t\geqslant0,i\in E,\\&p_{i\theta}(t)=1-\sum_{j\in E}p_{ij}(t), && t\geqslant0,i\in E,\end{aligned}$$

对新的状态空间 $\tilde{E}=E\cup\{\theta\}$,不难验证,$\{p_{ij}(t),t\geqslant0,i,j\in\tilde{E}\}$ 就是转移概率函数了.今后我们会解释为什么需要引进广转移概率函数.

对连续时间马尔可夫链 $X=\{X_t,t\geqslant0\}$,任取 $h>0$,定义

$$X_n(h)=X_{nh},\qquad n\geqslant0.$$

由马尔可夫性(1)立即知道,$\{X_n(h),n\geqslant0\}$ 是一个离散时间的马尔可夫链,称它为以 h 为步长的*离散骨架*,或简称 *h 骨架*. h 骨架的转移概率矩阵即为 $\mathbf{P}(h)=(p_{ij}(h))$,$n$ 步转移概率矩阵则为 $\mathbf{P}(nh)$.离散时间马尔可夫链的许多性质与结果,通过离散骨架必定在连续时间马尔可夫链身上有所反映.这也是我们利用已知结果讨论连续时间马尔可夫链的一条有效途径.

例 1.1 独立平稳增量过程 设 $\{X_t,t\geqslant0\}$ 为只取整数值的独立增量过程.与离散时间情形完全同样地可以证明,独立增量过程是马尔可夫链.若它有平稳增量:

$$\mathsf{P}(X_{t+s}-X_s=i)=p_i(t),\quad t\geqslant0,i\in\mathbb{Z},$$

则马尔可夫链是齐次的,其转移概率函数为

$$p_{ij}(t)=p_{j-i}(t), \qquad t\geqslant 0.$$

与离散时间情形完全类似地可以证明，若马尔可夫链的转移概率函数 $p_{ij}(t)$ 只依赖于 $j-i$（也称为空间齐次的），则它是独立平稳增量过程.

参数为 λ 的泊松过程为马尔可夫链，其转移概率函数为

$$p_{ij}(t)=\begin{cases}\dfrac{(\lambda t)^{j-i}}{(j-i)!}\mathrm{e}^{-\lambda t}, & j\geqslant i,\\ 0, & j<i.\end{cases}$$

这转移概率函数也称为泊松转移概率函数. 泊松过程也就是具有泊松转移概率函数且初值为零的马尔可夫链. □

下面我们总是对一个固定的连续时间齐次马尔可夫链进行讨论，不再重复指明.

定理 1.1 下列条件等价：

(1) $\lim\limits_{t\to 0^+}p_{ij}(t)=\delta_{ij}, \qquad i,j\in E;$ (9)

(2) 对任意的 $\varepsilon>0$ 及 $i\in E, t\geqslant 0$,

$$\lim_{h\to 0^+}\mathsf{P}_i(|X_{t+h}-X_t|\geqslant\varepsilon)=0. \tag{10}$$

（这里同样以 $\mathsf{P}_i(A)$ 记在 $X_0=i$ 的条件下事件 A 的条件概率 $\mathsf{P}(A\mid X_0=i)$.）

证 (1)⇒(2). 不妨设 $0<\varepsilon<1$，则 $h>0$ 时，

$$\begin{aligned}\mathsf{P}_i(|X_{t+h}-X_t|\geqslant\varepsilon)&=\mathsf{P}_i(X_{t+h}\neq X_t)\\&=\sum_j p_{ij}(t)\mathsf{P}(X_{t+h}\neq j\mid X_t=j)=\sum_j p_{ij}(t)(1-p_{jj}(h)).\end{aligned}$$

由于 $\lim_{h\to 0^+}(1-p_{jj}(h))=0$，且上式中的级数关于 h 一致收敛，在上式中令 $h\to 0^+$ 即得(10)式.

(2)⇒(1). 同样设 $0<\varepsilon<1$，在(10)式中取 $t=0$，则

$$1-p_{ii}(h)=\mathsf{P}_i(X_h\neq i)=\mathsf{P}_i(|X_h-X_0|\geqslant\varepsilon)\to 0, \qquad h\to 0^+.$$

对 $j\neq i$,

$$0\leqslant p_{ij}(h)\leqslant 1-p_{ii}(h)\to 0, \qquad h\to 0^+,$$

即(9)式成立. □

今后我们总假定，转移概率函数满足连续性条件(9)，即假设 $p_{ij}(t)$ 在 $t=0$ 点连续. 满足条件(9)的转移概率函数（或广转移概率函数）称为标准的. 从定理 1.1 的证明中可见，条件(9)等价于对一切 $i\in E$,

$$\lim_{t\to 0^+}p_{ii}(t)=1.$$

由定理 1.1 中的(10)式可知，连续性条件(9)相当于要求在将来很短时间内状态发生变化的概率也很小. 从实际应用方面来看，这样的要求也是自然的. 所以假定转移概率函数是标准的并不是严格的限制.

定理 1.2 对一切 $i\in E$ 及 $h>0,t\geqslant 0$,

$$\sum_j |p_{ij}(t+h)-p_{ij}(t)|\leqslant 2(1-p_{ii}(h)). \tag{11}$$

特别,对一切 $i,j\in E$,$p_{ij}(t)$为$[0,\infty)$上的一致连续函数.

证 由科尔莫戈罗夫-查普曼方程

$$\begin{aligned}p_{ij}(t+h)-p_{ij}(t)&=\sum_k p_{ik}(h)p_{kj}(t)-p_{ij}(t)\\&=\sum_{k\neq i}p_{ik}(h)p_{kj}(t)-(1-p_{ii}(h))p_{ij}(t),\end{aligned}$$

从而有

$$-(1-p_{ii}(h))p_{ij}(t)\leqslant p_{ij}(t+h)-p_{ij}(t)\leqslant \sum_{k\neq i}p_{ik}(h)p_{kj}(t),$$

$$\sum_j (p_{ij}(t+h)-p_{ij}(t))^+\leqslant \sum_j\sum_{k\neq i}p_{ik}(h)p_{kj}(t)=\sum_{k\neq i}p_{ik}(h)=1-p_{ii}(h),$$

$$\sum_j (p_{ij}(t+h)-p_{ij}(t))^-\leqslant \sum_j (1-p_{ii}(h))p_{ij}(t)=1-p_{ii}(h).$$

这里 $a^+=\max(a,0),a^-=\max(-a,0)$分别为实数 a 的正部及负部,而$|a|=a^++a^-$,由此即得(11)式. □

定理 1.2 表明,由于转移概率函数满足科尔莫戈罗夫-查普曼方程,在零点的连续性保证了它在每一点的连续性.事实上,我们将看到转移概率函数的性状,在很大程度上由它在零点附近的性状所决定,不仅仅对于连续性是这样.

定理 1.3 对任意的 $i\in E$,有 $p_{ii}(t)>0$.若有 $t_0>0$ 使 $p_{ij}(t_0)>0$,则 $t\geqslant t_0$ 时,$p_{ij}(t)>0$.

证 由科尔莫戈罗夫-查普曼方程,$u>0,v>0$ 时

$$p_{ii}(u+v)=\sum_k p_{ik}(u)p_{ki}(v)\geqslant p_{ii}(u)p_{ii}(v)$$

(这类不等式今后将经常使用),从而对任意的 $t>0$

$$p_{ii}(t)\geqslant\left[p_{ii}\left(\frac{t}{n}\right)\right]^n.$$

但 $\lim_{n\to\infty}p_{ii}(t/n)=1$,$n$ 充分大时 $p_{ii}(t/n)>0$,故 $p_{ii}(t)>0$.

设 $t_0>0,p_{ij}(t_0)>0$,则 $t>t_0$ 时

$$p_{ij}(t)\geqslant p_{ij}(t_0)p_{jj}(t-t_0)>0.\quad □$$

事实上,对任意的 $i,j\in E$,$p_{ij}(t)$在$(0,\infty)$上或恒为零,或恒为正.但这结论的证明比较冗长,我们不在这里讲述了.

引理 1.1 设定义在$[0,\infty)$上的连续函数 $g(t)$满足条件:

$$g(t+s)\leqslant g(t)+g(s),\qquad t,s>0 \tag{12}$$

及 $g(0)=0$,则

$$\lim_{t\to 0^+}\frac{g(t)}{t}=\sup_{t>0}\frac{g(t)}{t}.$$

(极限可能是$+\infty$.)

证 设$0<h<t$,$t=nh+s$,n为正整数,$s\in(0,h)$.由(12)式

$$\frac{g(t)}{t}\leqslant\frac{ng(h)}{t}+\frac{g(s)}{t}=\frac{g(h)}{h}\frac{nh}{t}+\frac{g(s)}{t},$$

令$h\to0$,则$s\to0$,因此

$$\frac{g(t)}{t}\leqslant\varliminf_{h\to0^+}\frac{g(h)}{h},$$

从而有

$$\varlimsup_{t\to0^+}\frac{g(t)}{t}\leqslant\sup_{t>0}\frac{g(t)}{t}\leqslant\varliminf_{h\to0^+}\frac{g(h)}{h},$$

由此即得$\lim_{t\to0^+}\frac{g(t)}{t}$存在且等于$\sup_{t>0}\frac{g(t)}{t}$. □

定理 1.4 对一切$i\in E$,存在极限

$$\lim_{t\to0^+}\frac{1-p_{ii}(t)}{t}=\sup_{t>0}\frac{1-p_{ii}(t)}{t}=q_i,\quad 0\leqslant q_i\leqslant\infty. \tag{13}$$

证 由定理1.3,$p_{ii}(t)>0$,$t\geqslant0$,定义$g(t)=-\ln p_{ii}(t)$,则$g(t)$为$[0,\infty)$上连续函数,$g(0)=0$,且满足(12)式,故存在极限

$$q_i=\lim_{t\to0^+}\frac{g(t)}{t}=\sup_{t>0}\frac{g(t)}{t}.$$

注意到$\lim_{t\to0^+}g(t)=0$,我们有

$$\lim_{t\to0^+}\frac{1-p_{ii}(t)}{t}=\lim_{t\to0^+}\frac{1-\mathrm{e}^{-g(t)}}{t}=\lim_{t\to0^+}\frac{g(t)}{t}\frac{1-\mathrm{e}^{-g(t)}}{g(t)}=q_i.$$

另一方面,对任意$t>0$,

$$p_{ii}(t)=\mathrm{e}^{-\frac{g(t)}{t}t}\geqslant\mathrm{e}^{-q_it},$$

$$\frac{1-p_{ii}(t)}{t}\leqslant\frac{1-\mathrm{e}^{-q_it}}{t}\leqslant q_i.$$

由此即得(13)式. □

定理 1.5 对任意的$i\neq j$,存在极限

$$\lim_{t\to0^+}\frac{p_{ij}(t)}{t}=q_{ij}<\infty. \tag{14}$$

证 对任意的$\varepsilon>0$,存在$\delta>0$,使$t\in(0,\delta)$时

$$1-p_{ii}(t)<\varepsilon,\qquad 1-p_{jj}(t)<\varepsilon.$$

先取$t\in(0,\delta)$,再取更小的$h\in(0,t)$,并设$nh\leqslant t<(n+1)h$,n为正整数.对$1\leqslant k\leqslant n$,定义

$$p_k = \mathrm{P}_i(X_{kh}=j, X_{vh}\neq j, 0<v<k),$$
$$r_k = \mathrm{P}_i(X_{kh}=i, X_{vh}\neq j, 0<v<k).$$

(这里实际上我们用到了链的 h 骨架.) $k\geqslant 2$ 时

$$p_k \geqslant \mathrm{P}_i(X_{kh}=j, X_{(k-1)h}=i, X_{vh}\neq j, 0<v<k-1)\geqslant r_{k-1}p_{ij}(h). \tag{15}$$

如定义 $r_0=1$,则(15)式对 $k=1$ 也成立.用首次进入法,

$$p_{ij}(t)\geqslant \sum_{k=1}^{n} p_k p_{jj}(t-kh)\geqslant (1-\varepsilon)\sum_{k=1}^{n} p_k \tag{16}$$

(因 $t-kh\in(0,\delta)$,所以 $p_{jj}(t-kh)>1-\varepsilon$).另一方面,

$$p_{ij}(t)\leqslant \sum_{k\neq i} p_{ik}(t) = 1-p_{ii}(t)<\varepsilon,$$

因此有 $\sum_{k=1}^{n} p_k \leqslant \frac{\varepsilon}{1-\varepsilon}$.在 $1\leqslant k\leqslant n$ 时,仍由首次进入法有

$$1-\varepsilon < p_{ii}(kh) = \sum_{v=1}^{k-1} p_v p_{ji}((k-v)h) + r_k$$
$$\leqslant \sum_{v=1}^{k-1} p_k + r_k \leqslant \frac{\varepsilon}{1-\varepsilon} + r_k,$$
$$r_k \geqslant 1-\varepsilon-\frac{\varepsilon}{1-\varepsilon}\geqslant \frac{1-3\varepsilon}{1-\varepsilon}. \tag{17}$$

由(15),(16)及(17)式,

$$p_{ij}(t)\geqslant \sum_{k=1}^{n} p_k p_{jj}(t-kh)\geqslant \sum_{k=1}^{n} r_{k-1}p_{ij}(h)p_{jj}(t-kh)$$
$$\geqslant n\frac{1-3\varepsilon}{1-\varepsilon}p_{ij}(h)(1-\varepsilon) = n(1-3\varepsilon)p_{ij}(h),$$
$$\frac{p_{ij}(h)}{h}\leqslant \frac{p_{ij}(t)}{nh}\frac{1}{1-3\varepsilon}\leqslant \frac{p_{ij}(t)}{t-h}\frac{1}{1-3\varepsilon},$$

令 $h\to 0^+$ 得

$$\varlimsup_{h\to 0^+}\frac{p_{ij}(h)}{h}\leqslant \frac{p_{ij}(t)}{t}\frac{1}{1-3\varepsilon}<\infty,$$

再令 $t\to 0^+$

$$\varlimsup_{h\to 0^+}\frac{p_{ij}(h)}{h}\leqslant \varliminf_{t\to 0^+}\frac{p_{ij}(t)}{t}\frac{1}{1-3\varepsilon},$$

最后令 $\varepsilon\to 0$ 即得(14)式. □

由于

$$\sum_{j\neq i}\frac{p_{ij}(t)}{t} = \frac{1-p_{ii}(t)}{t}\leqslant q_i, \qquad t>0,$$

令 $t\to 0^+$ 我们得到

$$\sum_{j\neq i} q_{ij} \leqslant q_i, \qquad i\in E. \tag{18}$$

定义 令 $q_{ii}=-q_i$，矩阵 $Q=(q_{ij})$ 称为马尔可夫链的密度矩阵或无穷小矩阵.若对一切 $i\in E$，

$$\sum_{j\neq i} q_{ij} = q_i < \infty, \tag{19}$$

则称密度矩阵或马尔可夫链是保守的.

易见，有限马尔可夫链总是保守的，因为在等式

$$\sum_{j\neq i} \frac{p_{ij}(t)}{t} = \frac{1-p_{ii}(t)}{t}$$

中取极限 $t\to 0^+$，直接由定理 1.5 得到(19)式.

我们指出，定理 1.4 及定理 1.5 以及(18)式对标准广转移概率函数也是成立的.因为只要把状态空间扩大，就可得到标准转移概率函数，然后就可应用定理 1.4 及定理 1.5 了.

从数学上看，q_{ij} 只是转移概率函数 $p_{ij}(t)$ 在零点的导数(所以，又称 Q 为 $\{\mathrm{P}(t), t\geqslant 0\}$ 的无穷小矩阵)，然而它们有着明确的概率意义.先阐述 q_i 的概率意义.若 $q_i=0$，由(13)式，这等价于对一切 $t\geqslant 0$，$p_{ii}(t)=1$.这意味着，系统一旦进入状态 i，就永远不再离开了.很自然地，我们称这样的状态为吸收状态.对一般的状态 i 及任意固定的 $t>0$，$n\to\infty$ 时，$p_{ii}(t/n)\to 1$，因此

$$2^n \ln p_{ii}\left(\frac{t}{2^n}\right) \simeq -2^n\left[1-p_{ii}\left(\frac{t}{2^n}\right)\right] \to -q_i t,$$

$$\mathrm{P}_i(X_{kt/2^n}=i, 1\leqslant k\leqslant 2^n) = \left[p_{ii}\left(\frac{t}{2^n}\right)\right]^{2^n} \to \mathrm{e}^{-q_i t}.$$

由连续性条件，我们可以期待 $n\to\infty$ 时应有

$$\mathrm{P}_i(X_{kt/2^n}=i, 1\leqslant k\leqslant 2^n) \to \mathrm{P}_i(X_s=i, s\in[0,t]).$$

(这不能算作严格的数学证明，只是直观的解释.但这并不妨碍我们理解密度矩阵的概率意义.)因此，我们有

$$\mathrm{P}_i(X_s=i, s\in[0,t]) = \mathrm{e}^{-q_i t}.$$

若 $q_i=\infty$，则过程不可能在状态 i 上停留一段时间，不管这段时间是多么地短，或者说，一进入状态 i 就必须立即离开它.所以把这样的状态称为瞬时状态.从理论上讲，这种情况是完全可能发生的，甚至可以人为地构造出这样一个马尔可夫链，其中每一个状态都是瞬时状态.然而从实际应用的角度来看，这种状态是难以理解的.也就是说，在实际应用中遇到的马尔可夫链一般是不会发生这种情况的.所以我们今后不讨论这种状态，即假设对一切 $i\in E$，$q_i<\infty$.现在令

$$\tau = \inf\{t>0: X_t\neq X_0\}.$$

在 $X_0=i$ 的条件下，τ 就是在状态 i 停留的时间.由前面的讨论，我们有

$$\mathrm{P}_i(\tau>t) = \mathrm{P}_i(X_s=i, s\in[0,t]) = \mathrm{e}^{-q_i t}.$$

在 $0<q_i<\infty$ 时,在状态 i 上的停留时间服从参数为 q_i 的指数分布,平均停留时间为 $1/q_i$. 停留时间服从指数分布显然是由马尔可夫性所决定的. 我们知道,具有无后效性的唯一的连续型分布是指数分布. q_i 的概率意义也提供了计算 q_i 的方法. 我们只要计算当系统处于状态 i 时,在短时间 t 之后离开状态 i 的概率的主要线性部分:

$$1-p_{ii}(t)=q_it+o(t),\qquad t\to 0.$$

我们再解释 $q_{ij}, i\neq j$ 的概率意义. 设 i 不是吸收状态. 当系统处于状态 i 时,在短时间 t 之后离开状态 i 的条件下进入状态 j 的条件概率为

$$\frac{p_{ij}(t)}{1-p_{ii}(t)}\to\frac{q_{ij}}{q_i},\quad t\to 0.$$

因此,q_{ij}/q_i 可解释为,在系统要离开原来的状态 i 的条件下,进入状态 j 的条件概率. 从这个意义出发,自然应要求 $\sum_{j\neq i}q_{ij}/q_i=1$,这就是保守性的要求. 因此可以预料,在实际应用中得到的马尔可夫链都是保守的. 在本课程中也将只讨论保守马尔可夫链. 若 i 是吸收状态:$q_i=0$,由(19),对一切 $j\neq i$,$q_{ij}=0$.

定义 对保守马尔可夫链,称随机矩阵

$$\mathbf{R}=(r_{ij}),\quad r_{ij}=\begin{cases}(1-\delta_{ij})\dfrac{q_{ij}}{q_i}, & 若\ q_i>0,\\ \delta_{ij}, & 若\ q_i=0\end{cases}$$

为跳跃阵. 以 $\mathbf{R}$ 为转移概率矩阵的离散时间马尔可夫链称为原来的连续时间马尔可夫链的跳跃链.

跳跃链刻画了原来的链的状态转移的状况,是研究连续时间马尔可夫链的一个重要手段. 为了计算 q_{ij},我们也只要计算从状态 i 出发经过短时间 t 到达状态 j 的概率的主要线性部分:

$$p_{ij}(t)=q_{ij}t+o(t),\qquad t\to 0.$$

例 1.2 生灭过程 一个细菌繁殖的模型如下. 一个细菌在时间区间 $(t,t+\Delta t)$ 中分裂为两个细菌的概率为 $\lambda\Delta t+o(\Delta t)$,死亡的概率为 $\mu\Delta t+o(\Delta t)$,既不分裂也不死亡的概率为 $1-(\lambda+\mu)\Delta t+o(\Delta t)$,它们都与这细菌在时刻 t 之前已存活多久没有关系. 各个细菌的变化状况假定是相互独立的. 以 X_t 表示时刻 t 时细菌的总数,则 $\{X_t,t\geqslant 0\}$ 是马尔可夫链. 我们来计算它的密度矩阵. 显然,0 是吸收状态,细菌都死亡了,过程也就停止了. 所以 $p_{00}(t)\equiv 1$, $q_{0i}=0$, $i\geqslant 0$. 设 $X_t=i\neq 0$. 由独立性的假设可知,在 $(t,t+\Delta t)$ 中有两个或两个以上的细菌发生变化(分裂或死亡)的概率是 $o(\Delta t)$,因为一个细菌发生变化的概率是 Δt 的一阶无穷小,两个或两上以上的一阶无穷小相乘至少是两阶无穷小. 我们只要考虑有一个细菌发生变化而其它细菌不变,或全部细菌都不变化的情形:

$$\begin{aligned}&p_{ij}(\Delta t)=o(\Delta t),\qquad |i-j|\geqslant 2,\\&p_{i,i+1}(\Delta t)=i\lambda\Delta t+o(\Delta t),\\&p_{i,i-1}(\Delta t)=i\mu\Delta t+o(\Delta t),\\&p_{ii}(\Delta t)=[1-(\lambda+\mu)\Delta t+o(\Delta t)]^i=1-i(\lambda+\mu)\Delta t+o(\Delta t).\end{aligned}$$

这里只要注意，增加(减少)一个细菌可以是 i 个细菌中的任意一个分裂(死亡)，因而 Δt 的系数为 $i\lambda(i\mu)$. 现在即可写出密度矩阵为

$$Q=\begin{pmatrix}0 & 0 & 0 & 0 & 0 & \cdots\\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots\\ 0 & 2\mu & -2(\lambda+\mu) & 2\lambda & 0 & \cdots\\ 0 & 0 & 3\mu & -3(\lambda+\mu) & 3\lambda & \cdots\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\end{pmatrix},$$

跳跃阵为

$$\mathbf{R}=\begin{pmatrix}1 & 0 & 0 & 0 & \cdots\\ q & 0 & p & 0 & \cdots\\ 0 & q & 0 & p & \cdots\\ \cdots & \cdots & \cdots & \cdots & \cdots\end{pmatrix}, p=\frac{\lambda}{\lambda+\mu}, q=\frac{\mu}{\lambda+\mu}.$$

一般地，若一个马尔可夫链的密度矩阵有下列形式：

$$Q=\begin{pmatrix}-\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots\\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & 0 & \cdots\\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & 0 & \cdots\\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \lambda_3 & \cdots\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\end{pmatrix},$$

其中 $\lambda_i\geqslant 0, i\geqslant 0; \mu_i\geqslant 0, i\geqslant 1; \lambda_i+\mu_i>0, i\geqslant 1$，则称它为生灭过程，这密度矩阵也称为生灭矩阵. λ_i 称为出生率，μ_i 称为死亡率. 前面所述细菌繁殖的模型称为线性生灭过程，因为这时出生率 $\lambda_i=i\lambda$ 与死亡率 $\mu_i=i\mu$ 均为 i 的线性函数. 生灭过程的状态转移只能是从 i 到 $i+1$，看作为出生一个，或从 i 到 $i-1$，看作为死亡一个. 这就是生灭过程的名称的由来. 这样的状态转移与离散时间情形的随机游动是相似的. 实际上，生灭过程的跳跃链就是随机游动，跳跃阵为

$$\mathbf{R}=\begin{pmatrix}r_0 & p_0 & 0 & 0 & \cdots\\ q_1 & 0 & p_1 & 0 & \cdots\\ 0 & q_2 & 0 & p_2 & \cdots\\ \cdots & \cdots & \cdots & \cdots & \cdots\end{pmatrix},$$

其中 $p_i=\frac{\lambda_i}{\lambda_i+\mu_i}, q_i=\frac{\mu_i}{\lambda_i+\mu_i}, i\geqslant 1$；当 $\lambda_0=0$ 时 $r_0=1, p_0=0$；$\lambda_0>0$ 时 $r_0=0, p_0=1$. 所以生灭过程与离散时间的随机游动在性质上有许多相似之处.

如果一个生灭过程满足条件 $\mu_i=0, i\geqslant 1$，即只出生而无死亡，则称它为纯生过程. 泊松过程就是出生率为常数的纯生过程. 类似地，若 $\lambda_i=0, i\geqslant 0$，即只死亡而无出生，则称它为纯灭过程.

生灭过程是密度矩阵的形式最简单的一类马尔可夫链，也是在实际应用中大量遇到的一类随机过程，因此无论在理论上还是在应用上，它都显得十分重要. 此外，正如离散时间的随机游动的状态空间可以是 $\mathbb{Z}=\{\cdots,-1,0,1,\cdots\}$，生灭过程也可以 $\mathbb{Z}$ 为状态空间，这时称之为双边生灭

过程. □

在实际应用中,如例 1.2 中所示的那样,一般是根据实际情况或理论上的假设来断定,一个系统的状态的变化是一个连续时间的马尔可夫链.然而要直接写出转移概率函数一般是难以做到的.这与离散时间的马尔可夫链在实际应用中的情况很不相同.在离散时间情形,一般都能写出转移概率矩阵,然后在此基础上作进一步的讨论.但是在连续时间情形,根据对具体情况的分析与假定,往往可以写出密度矩阵(在很大程度上可以说,连续时间情形中的密度矩阵与离散时间情形中的转移概率矩阵地位相当).因此怎样由密度矩阵确定转移概率函数,就成为一个具有重大意义的问题.它不仅在理论上占有头等重要的位置,而且也是实际应用所迫切需要解决的问题.这就是马尔可夫链理论中著名的 Q 过程问题.

定义 一个矩阵 $Q=(q_{ij})$ 称为一个 Q 矩阵,若

$$q_{ij}\geqslant 0,\quad i\neq j,\quad -\infty<q_{ii}\leqslant 0,$$
$$\sum_{j\neq i}q_{ij}\leqslant q_{ii},\qquad i,j\in E.$$

设给定一个 Q 矩阵.若存在一个转移概率函数 $\{\mathbf{P}(t),t\geqslant 0\}$ 以 Q 为密度矩阵,则称 $\{\mathbf{P}(t),t\geqslant 0\}$ 为一个 Q 过程,或 Q 过程问题的一个解.

实际上,一般还把对 Q 过程的要求放宽为广义转移概率函数(密度矩阵同样地定义),因为从理论研究的角度来看,讨论广义转移概率函数更自然.所谓 Q 过程问题是指:(1) 解是否存在(存在性问题);(2) 解如果存在是否唯一(唯一性问题);(3) 如果解不唯一,找出全部解(构造问题).我国的数学家在 Q 过程问题的研究中做出了突出的贡献.详细地讨论 Q 过程问题必须进入专门的研究领域.

习　题

1-1. 证明:对任意的 $h>0$ 及 $i\in E$, $\sum_i |p_{ij}(t+h)-p_{ij}(t)|$ 是 t 的单调下降函数.

1-2. 设马尔可夫链是有限状态的.证明:矩阵 $\mathbf{P}(t)$ 的行列式 $\det\mathbf{P}(t)>0,t\geqslant 0$.

1-3. 设 $(p_{ij}(t))$ 为广转移概率函数.定义

$$d_i(t)=1-\sum_j p_{ij}(t),\qquad i\in E,t\geqslant 0.$$

证明:(1) $d_i(t)$ 为单调增加函数;

(2) 若存在 $t_0>0$ 使对一切 $i\in E,d_i(t_0)=0$,则对一切 $t>0$ 及 $i\in E,d_i(t)=0$.

1-4. 证明:对一切 $h>0$ 及 $i,j\in E$

$$|p_{ij}(t+h)-p_{ij}(t)|\leqslant q_ih.$$

1-5. 设出租汽车按参数为 λ 的泊松过程到达一出租汽车站候客,顾客按参数为 μ 的泊松过程来到车站候车,这两个泊松过程相互独立.假设一辆车载一位顾客,顾客到站时有空车即乘车离去,无空车也离站而去不等车.以 X_t 记时刻 t 车站上的出租汽车数,则 $\{X_t,t\geqslant 0\}$ 是马尔可夫链,写出它的密度矩阵.

1-6. 设顾客按参数为 λ 的泊松过程来到一理发店,且设在初始时刻 $t=0$ 已有一位顾客.每个顾客相互独立地以概率 p 为男性,以概率 $1-p$ 为女性.令

$$X_t=\begin{cases}0, & \text{若到时刻 } t \text{ 为止最后来到的顾客是男性,}\\ 1, & \text{若到时刻 } t \text{ 为止最后来到的顾客是女性,}\end{cases}\qquad t\geqslant 0,$$

则 $\{X_t, t\geqslant 0\}$ 是马尔可夫链,写出它的密度矩阵.

1-7. 设某生物群体中一个个体的寿命服从参数为 λ 的指数分布,一个个体死亡时有 k 个后代接续的概率为 $g_k, k=0,1,\cdots$,各个个体的后代状况相互独立. 以 X_t 记时刻 t 群体中个体的总数,则 $\{X_t, t\geqslant 0\}$ 是马尔可夫链,写出它的密度矩阵.

1-8. 设 Q 为有限马尔可夫链的密度矩阵,则

(1) 0 为 Q 的特征值;

(2) Q 的非零特征值有负实部.

§3.2 科尔莫戈罗夫方程

在上一节中,我们引进了密度矩阵,实际上是讨论了转移概率函数在 $t=0$ 的可微性问题. 由于我们假设 $q_i<\infty, i\in E$,定理 1.4 及 1.5 表示 $p_{ij}(t)$ 在 $t=0$ 点有(右)导数. 实际上,这也保证了 $p_{ij}(t)$ 在每一点可微. 但为了使问题简化,我们在适当条件下,证明转移概率函数可微,且得到它们满足的微分方程.

引理 2.1 设 $f(x)$ 为 (a,b) 上的连续函数,且 $f(x)$ 在 (a,b) 中有连续的右导数,则 $f(x)$ 在 (a,b) 上可导.

证 设 $f(x)$ 在 (a,b) 中有连续的右导数 $f_+(x)$. 任取 $x_0\in(a,b)$,令

$$g(x)=\int_{x_0}^{x} f_+(y)\mathrm{d}y,\qquad x\in(a,b),$$

则 $g(x)$ 在 (a,b) 上连续可导,且 $g'(x)=f_+(x)$.

$F(x)=f(x)-g(x)$ 在 (a,b) 上连续,有恒为零的右导数,我们只要证明 $F(x)$ 在 (a,b) 上为常数,从而 $F'(x)=0$, $f(x)$ 在 (a,b) 上可导,且 $f'(x)=g'(x)=f_+(x)$. 下面就假设 $f(x)$ 在 (a,b) 上有恒为零的右导数,我们要证 $f(x)$ 在 (a,b) 上为常数.

任取 $x_0\in(a,b)$ 及 $\varepsilon>0$,令

$$c=\sup\{x:\ \ x_0<x<b,\ \ |f(x)-f(x_0)|<\varepsilon(x-x_0)\}.$$

由于 $\lim_{x\to x_0}\dfrac{f(x)-f(x_0)}{x-x_0}=0$,必有 $\delta>0$,使 $x\in(x_0,x_0+\delta)$ 时 $|f(x)-f(x_0)|<\varepsilon(x-x_0)$,因此 $x_0<c\leqslant b$. 若 $c<b$,由 $f(x)$ 的连续性及 $x\in(x_0,c)$ 时 $|f(x)-f(x_0)|<\varepsilon(x-x_0)$ 可知

$$|f(c)-f(x_0)|\leqslant\varepsilon(c-x_0).$$

仍由 $\lim_{x\to c^+}\dfrac{f(x)-f(c)}{x-c}=0$,存在 $\delta'>0$,使 $x\in(c,c+\delta')$ 时

$$|f(x)-f(c)|<\varepsilon(x-c),$$

从而 $x\in(c,c+\delta')$ 时

$$|f(x)-f(x_0)|\leqslant|f(x)-f(c)|+|f(c)-f(x_0)|$$

$$<\varepsilon(x-x_0).$$

这与 c 的定义相矛盾,因此 $c=b$.这表明,$x>x_0$ 时总有 $|f(x)-f(x_0)|<\varepsilon(x-x_0)$.但 ε 可任意地小,因此 $x>x_0$ 时 $f(x)=f(x_0)$.这即说明 $f(x)$在(a,b)上为常数. □

定理 2.1 设链为保守的,则对一切 $i,j\in E$ 及 $t\geqslant 0$ 成立

$$p_{ij}'(t)=\sum_k q_{ik}p_{kj}(t). \tag{1}$$

证 由科尔莫戈罗夫-查普曼方程,$h>0$ 时

$$\frac{p_{ij}(t+h)-p_{ij}(t)}{h}=\frac{p_{ii}(h)-1}{h}p_{ij}(t)+\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t). \tag{2}$$

由上一章的引理 6.1,

$$\varliminf_{h\to 0^+}\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t)\geqslant\sum_{k\neq i}q_{ik}p_{kj}(t). \tag{3}$$

另一方面,对 $N>i$

$$\begin{aligned}\varlimsup_{h\to 0^+}\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t)&\leqslant\varlimsup_{h\to 0^+}\left[\sum_{k\neq i,k<N}\frac{p_{ik}(h)}{h}p_{kj}(t)+\sum_{k\geqslant N}\frac{p_{ik}(h)}{h}\right]\\&=\varlimsup_{h\to 0^+}\left[\sum_{k\neq i,k<N}\frac{p_{ik}(h)}{h}p_{kj}(t)+\frac{1-p_{ii}(h)}{h}-\sum_{k\neq i,k<N}\frac{p_{ik}(h)}{h}\right]\\&=\sum_{k\neq i,k<N}q_{ik}p_{kj}(t)+q_i-\sum_{k\neq i,k<N}q_{ik},\end{aligned}$$

令 $N\to\infty$,由保守性得

$$\varlimsup_{h\to 0^+}\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t)\leqslant\sum_{k\neq i}q_{ik}p_{kj}(t). \tag{4}$$

由(3)及(4)得

$$\lim_{h\to 0^+}\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t)=\sum_{k\neq i}q_{ik}p_{kj}(t).$$

再由(2)得

$$\lim_{h\to 0^+}\frac{p_{ij}(t+h)-p_{ij}(t)}{h}=\sum_k q_{ik}p_{kj}(t).$$

仍由保守性知,上式右边的级数关于 t 一致收敛,因此是 t 的连续函数,故由引理 2.1 得(1). □

定理 2.2 设链为保守的,且 $q=\sup_i q_i<\infty$(满足这条件的密度矩阵称为有界的),则对一切 $i,j\in E$ 及 $t\geqslant 0$ 成立

$$p_{ij}'(t)=\sum_k p_{ik}(t)q_{kj}. \tag{5}$$

证 由于链是保守的(实际上密度矩阵有界的链必定是保守的,我们不在这里讨论这一点),

因此由定理 2.1,转移概率函数的导数存在. 设 $h>0, k\neq j$,则

$$0\leqslant\frac{p_{kj}(h)}{h}\leqslant\sum_{l\neq k}\frac{p_{kl}(h)}{h}=\frac{1-p_{kk}(h)}{h}\leqslant q_k\leqslant q<\infty.$$

在等式

$$\frac{p_{ij}(t+h)-p_{ij}(t)}{h}=p_{ij}(t)\frac{p_{jj}(h)-1}{h}+\sum_{k\neq j}p_{ik}(t)\frac{p_{kj}(h)}{h}$$

中令 $h\to0^+$,注意到上式中的级数关于 h 一致收敛,即得

$$p'_{ij}(t)=p_{ij}(t)q_{jj}+\sum_{k\neq j}p_{ik}(t)q_{kj}.\quad\square$$

方程(1)及(5)称为科尔莫戈罗夫方程,其中(1)称为向后方程,(5)称为向前方程. 形式上对科尔莫戈罗夫－查普曼方程

$$p_{ij}(t+s)=\sum_k p_{ik}(t)p_{kj}(s)$$

求导就得到科尔莫戈罗夫方程. 对后面的参数 t 在零点求导,即得向后方程. 对前面的参数 s 在零点求导,即得向前方程. 这也就是这些名称的由来. 科尔莫戈罗夫方程的矩阵形式也是十分简洁的:

$$\mathbf{P}'(t)=\boldsymbol{Q}\mathbf{P}(t)\qquad(\text{向后方程}),$$
$$\mathbf{P}'(t)=\mathbf{P}(t)\boldsymbol{Q}\qquad(\text{向前方程}).$$

科尔莫戈罗夫方程提供了由密度矩阵求转移概率函数的一条途径:解微分方程组. 但这在理论上也产生一系列的问题,那就是科尔莫戈罗夫方程解的存在性及唯一性问题,以及在解不唯一的情况下如何找出全部解的问题. 从 Q *过程问题*的角度看也就是满足科尔莫戈罗夫方程的 Q 过程的存在性、唯一性及构造问题. 众所周知,要得到微分方程解的明显表达式,一般来说是极为困难的. 然而,尽管难以求得科尔莫戈罗夫方程解的表达式,我们仍可利用科尔莫戈罗夫方程去讨论转移概率函数的性质,特别可以利用它去判明在什么条件下转移概率函数为它的密度矩阵唯一决定,进而只要讨论密度矩阵的性质,就能掌握转移概率函数的性质. 在理论上讨论向后方程比较方便. 讨论向前方程比较困难. 比较一下定理 2.1 及定理 2.2 即可看到这一点. 向后方程成立的条件比较弱,只要链是保守的就行. 我们早已指出,在实际应用中遇到的马尔可夫链理应是保守的.(事实上,保守性也是向后方程成立的必要条件,但我们不去作深入讨论了.)然而我们给出的向前方程成立的条件要强得多. 但是在求解方程时,向前方程更为适用,因而它同样是十分重要的.

例 2.1　有限马尔可夫链　对有限马尔可夫链,直接在科尔莫戈罗夫－查普曼方程中求导,就得到科尔莫戈罗夫向前与向后方程:

$$\mathbf{P}'(t)=\mathbf{P}(t)\boldsymbol{Q}=\boldsymbol{Q}\mathbf{P}(t),\qquad t\geqslant0,$$

初始条件为 $\mathbf{P}(0)=\boldsymbol{I}$. 这常系数线性方程组的解是

$$\mathbf{P}(t)=\mathrm{e}^{Qt}=\sum_{n=0}^{\infty}\frac{(Qt)^n}{n!},\qquad t\geqslant0.$$

所以对有限马尔可夫链,其转移概率函数为它的密度矩阵唯一决定.科尔莫戈罗夫方程的作用在这里是显而易见的.显然,在状态空间为无穷时,问题将要复杂与困难得多. □

例 2.2 设密度矩阵为

$$Q=\begin{pmatrix}-2 & 1 & 1\\ 1 & -4 & 3\\ 2 & 1 & -3\end{pmatrix}.$$

我们来求转移概率函数$\{\mathbf{P}(t),t\geqslant 0\}$.为了便于用指数级数计算,先要将 Q 对角化.Q 的特征方程为:$\det(Q-\lambda I)=\lambda(\lambda+4)(\lambda+5)=0$,特征值为:$0,-4,-5$,相应的特征向量可取为:$(1,1,1),(-3,5,1),(1,-4,1)$.事实上,0 总是有限密度矩阵的特征值,相应的特征向量可取为$(1,1,1)$(见习题 1-8).因此 $Q=MJM^{-1}$,其中

$$M=\begin{pmatrix}1 & -3 & 1\\ 1 & 5 & -4\\ 1 & 1 & 1\end{pmatrix},\quad J=\begin{pmatrix}0 & 0 & 0\\ 0 & -4 & 0\\ 0 & 0 & -5\end{pmatrix},\quad M^{-1}=\frac{1}{20}\begin{pmatrix}9 & 4 & 7\\ -5 & 0 & 5\\ -4 & -4 & 8\end{pmatrix}.$$

所以 $t\geqslant 0$ 时,

$$\mathbf{P}(t)=\mathrm{e}^{Qt}=M\sum_{n=0}^{\infty}\frac{(Jt)^n}{n!}M^{-1}=M\begin{pmatrix}1 & 0 & 0\\ 0 & \mathrm{e}^{-4t} & 0\\ 0 & 0 & \mathrm{e}^{-5t}\end{pmatrix}M^{-1}$$

$$=\frac{1}{20}\begin{pmatrix}9+15\mathrm{e}^{-4t}-4\mathrm{e}^{-5t} & 4-4\mathrm{e}^{-5t} & 7-15\mathrm{e}^{-4t}+8\mathrm{e}^{-5t}\\ 9-25\mathrm{e}^{-4t}+16\mathrm{e}^{-5t} & 4+16\mathrm{e}^{-5t} & 7+25\mathrm{e}^{-4t}-32\mathrm{e}^{-5t}\\ 9-5\mathrm{e}^{-4t}-4\mathrm{e}^{-5t} & 4-4\mathrm{e}^{-5t} & 7+5\mathrm{e}^{-4t}+8\mathrm{e}^{-5t}\end{pmatrix}. \quad \square$$

众所周知,把微分方程写成积分方程的形式讨论起来更为方便.下面我们就给出科尔莫戈罗夫方程的积分形式.

向后方程的积分方程形式为:对一切 $i,j\in E$ 及 $t\geqslant 0$,

$$p_{ij}(t)=\delta_{ij}\mathrm{e}^{-q_it}+\sum_{k\neq i}\int_0^t \mathrm{e}^{-q_i(t-s)}q_{ik}p_{kj}(s)\mathrm{d}s. \tag{6}$$

实际上,由向后方程(1)

$$\begin{aligned}p_{ij}(t)-\delta_{ij}\mathrm{e}^{-q_it}&=\int_0^t \mathrm{d}[p_{ij}(s)\mathrm{e}^{-q_i(t-s)}]\\&=\int_0^t \mathrm{e}^{-q_i(t-s)}p'_{ij}(s)\mathrm{d}s+\int_0^t q_i\mathrm{e}^{-q_i(t-s)}p_{ij}(s)\mathrm{d}s\\&=\int_0^t \mathrm{e}^{-q_i(t-s)}\Big(q_{ii}p_{ij}(s)+\sum_{k\neq i}q_{ik}p_{kj}(s)\Big)\mathrm{d}s-\int_0^t q_{ii}\mathrm{e}^{-q_i(t-s)}p_{ij}(s)\mathrm{d}s\\&=\sum_{k\neq i}\int_0^t \mathrm{e}^{-q_i(t-s)}q_{ik}p_{kj}(s)\mathrm{d}s,\end{aligned}$$

即得(6).由于 $\sum_{k\neq i}q_{ik}p_{kj}(s)$关于 s 一致收敛,为 s 的连续函数,对(6)式求导即得向后方程(1).

向前方程(5)的积分方程形式为：对一切 $i,j\in E$ 及 $t\geqslant 0$

$$p_{ij}(t)=\delta_{ij}\mathrm{e}^{-q_j t}+\sum_{k\neq i}\int_0^t p_{ik}(s)q_{kj}\mathrm{e}^{-q_j(t-s)}\mathrm{d}s. \tag{7}$$

类似地，由向前方程(5)，

$$\begin{aligned}
p_{ij}(t)-\delta_{ij}\mathrm{e}^{-qi} &= \int_0^t \mathrm{d}[p_{ij}(s)\mathrm{e}^{-q_i(t-s)}] \\
&= \int_0^t p'_{ij}(s)\mathrm{e}^{-q_j(t-s)}\mathrm{d}s+\int_0^t p_{ij}(s)q_j\mathrm{e}^{-q_j(t-s)}\mathrm{d}s \\
&= \int_0^t \Big(p_{ij}(s)q_{jj}+\sum_{k\neq j}p_{ij}(s)q_{kj}\Big)\mathrm{e}^{-q_j(t-s)}\mathrm{d}s-\int_0^t p_{ij}(s)q_{jj}\mathrm{e}^{-q_j(t-s)}\mathrm{d}s \\
&= \sum_{k\neq j}\int_0^t p_{ik}(s)q_{kj}\mathrm{e}^{-q_j(t-s)}\mathrm{d}s,
\end{aligned}$$

即得(7)式. 如果 $\sum_{k\neq j}p_{ik}(s)q_{kj}$ 为连续函数，对(7)式求导即得向前方程(5).

将科尔莫戈罗夫方程写成积分形式(6)及(7)是有概率意义的. 先给出(6)式(也称为向后方程)的直观解释. 我们将从状态 i 出发经过时间 t 之后到达状态 j 这一事件做如下的分解：首先考虑在 $[0,t]$ 内始终没有离开状态 i 的情况，这时经过时间 t 由 i 到 j 的概率就是 $\delta_{ij}\mathrm{e}^{-q_i t}$ ($\mathrm{e}^{-q_i t}$ 是在状态 i 上停留时间超过 t 的概率). 其次，考虑在状态 i 上停留的时间 τ 小于 t 的情况. $\tau\in(t-s,t-s+\mathrm{d}s)$ 的概率为 $q_i\mathrm{e}^{-q_i(t-s)}\mathrm{d}s$，在时刻 τ 由状态 i 跳至状态 $k(\neq i)$ 的概率为 q_{ik}/q_i，然后由状态 k 经过时间 s 到达状态 j 的概率为 $p_{kj}(s)$，τ 的取值范围是 $(0,t)$，因此全部概率为

$$\sum_{k\neq i}\int_0^t q_i\mathrm{e}^{-q_i(t-s)}\frac{q_{ik}}{q_i}p_{kj}(s)\mathrm{d}s.$$

这就是(6)式右边的第二项($q_i=0$ 时 i 为吸收状态，(6)式即为 $p_{ij}(t)=\delta_{ij}$，第二项为 0).

对(7)式(也称为向前方程)可作类似的解释，但是代替上面考虑的在状态 i 上停留的时间，我们考虑在时刻 t 之前最后一次发生状态改变的时刻，然后作下面的分解：(7)式右边的第一项同样是在 $(0,t)$ 中没有发生状态转移情况下，经过时间 t 由 i 到 j 的概率. 现在考虑在时刻 t 到达 j 之前是由状态 $k(\neq j)$ 跳到 j 的情况. 记 σ_k 为时刻 t 之前最后入状态 k 的时刻，然后在状态 k 上停留的时间为 τ_k. 在初始状态为 i 的条件下，σ_k 的分布函数记为 $G_{ik}(u)$. τ_k 服从参数为 q_k 的指数分布，且与 σ_k 独立. 因此，由卷积公式可知，$\sigma_k+\tau_k$ 有分布密度 $\int_0^t q_k\mathrm{e}^{-q_k(t-u)}\mathrm{d}G_{ik}(u)$. 事实上，若计算 $p_{ik}(t)$，这时 $\tau_k\geqslant t-\sigma_k\geqslant 0$，从而 $p_{ik}(t)=\int_0^t \mathrm{e}^{-q_k(t-u)}\mathrm{d}G_{ik}(u)$. 换句话说，$\sigma_k+\tau_k$ 的分布密度就是 $p_{ik}(s)q_k$，$\sigma_k+\tau_k\in(s,s+\mathrm{d}s)$ 的概率为 $p_{ik}(s)q_k\mathrm{d}s$，接着由状态 k 跳到状态 j 的概率为 q_{kj}/q_k，然后在状态 j 上停留时间超过 $t-s$ 的概率为 $\mathrm{e}^{-q_j(t-s)}$，$\sigma_k+\tau_k$ 的取值范围为 $(0,t)$. 因此在 $(0,t)$ 中发生状态转移的情况下，经过时间 t 由 i 到 j 的全部概率为

$$\sum_{k\neq j}\int_0^t p_{ik}(s)q_k\frac{q_{kj}}{q_k}\mathrm{e}^{-q_j(t-s)}\mathrm{d}s.$$

这就是(7)式右边的第二项. 必须指出，上面只是直观地解释了科尔莫戈罗夫方程的积分形式的

概率意义,不能作为严格的数学证明.要使这些解释通得过,必须有一定的条件,这也就是使科尔莫戈罗夫方程成立的条件.

拉普拉斯(Laplace)变换是讨论微分方程或积分方程的一个有力工具,它将这些方程转化成代数方程,以便于用代数方法作处理,所以我们也要用到它.为此首先叙述一些必需的有关拉普拉斯变换的基本事实.

定义 设 $f(t)$为定义在$[0,\infty)$上的实函数.若对任意的 $\lambda>0$,积分

$$\phi(\lambda)=\int_0^\infty \mathrm{e}^{-\lambda t}f(t)\mathrm{d}t$$

存在,则称 $\phi(\lambda),\lambda>0$ 为 $f(t)$的**拉普拉斯变换**.显然,$[0,\infty)$上的有界函数总有拉普拉斯变换.易见,常数 $f(t)\equiv c$ 的拉普拉斯变换为c/λ;指数函数 $f(t)=\mathrm{e}^{-qt},q>0$ 的拉普拉斯变换为 $1/(\lambda+q)$.下列引理解决了拉普拉斯变换的唯一性问题.

引理 2.2 设$[0,\infty)$上的连续函数 $f(t)$有拉普拉斯变换 $\phi(\lambda)\equiv0,\lambda>0$,则 $f(t)\equiv0,t\geqslant0$

证 先设 $f(t)$在$[0,\infty)$上可积:$\int_0^\infty|f(t)|\mathrm{d}t<\infty$.作变换 $x=\mathrm{e}^{-t}$,则 $g(x)=f(-\ln x)/x$ 为$(0,1]$上的连续函数,且$\int_0^1|g(x)|\mathrm{d}x=\int_0^\infty|f(t)|\mathrm{d}t<\infty$.由假设条件,对一切 $n\geqslant1$,

$$\int_0^1 x^n g(x)\mathrm{d}x=\int_0^\infty \mathrm{e}^{-nt}f(t)\mathrm{d}t=0,$$

从而对一切多项式 $h(x)$

$$\int_0^1 h(x)g(x)\mathrm{d}x=h(0)\int_0^1 g(x)\mathrm{d}x. \tag{8}$$

由于$[0,1]$上的连续函数可用多项式一致逼近,因此(8)式对$[0,1]$上的任意连续函数 $h(x)$也成立.下证 $g(x)$在$(0,1]$上恒为零.我们可用反证法.设 $x_0\in(0,1]$使 $g(x_0)\neq0$,不妨设 $g(x_0)>0$.由 $g(x)$的连续性,存在 $\delta>0$,使 $x\in(x_0-\delta,x_0+\delta)$时 $g(x)>0$,且 $x_0-\delta>0$.取$[0,1]$上的连续函数 $h(x)$,使得 $h(x)>0,x\in(x_0-\delta,x_0+\delta)$,而 $h(x)=0,x\notin(x_0-\delta,x_0+\delta)$,特别 $h(0)=0$.由(8)式,$\int_0^1 h(x)g(x)\mathrm{d}x=0$,但$\int_0^1 h(x)g(x)\mathrm{d}x=\int_{x_0-\delta}^{x_0+\delta}h(x)g(x)\mathrm{d}x$ 应该大于零,得出矛盾.所以,$g(x)=0,x\in(0,1]$,从而 $f(t)=0,t\in[0,\infty)$.

现在去掉条件$\int_0^\infty|f(t)|\mathrm{d}t<\infty$.任取 $\delta>0$,令 $f_1(t)=\mathrm{e}^{-\delta t}f(t)$,则由于 f 的拉普拉斯变换存在,$\int_0^\infty|f_1(t)|\mathrm{d}t<\infty$,且 $f_1(t)$的拉普拉斯变换为 $\phi(\lambda+\delta)=0,\lambda>0$.由已证结果,$f_1(t)\equiv0$,从而仍有 $f(t)\equiv0$. □

引理 2.3 设$[0,\infty)$上的函数 $f(t)$的拉普拉斯变换为 $\phi(\lambda)$,则

$$g(t)=\int_0^t \mathrm{e}^{-q(t-s)}f(s)\mathrm{d}s,\quad q>0$$

的拉普拉斯变换为$\dfrac{\phi(\lambda)}{\lambda+q}$.

证 直接计算即可：

$$\int_0^\infty e^{-\lambda t}dt\int_0^t e^{-q(t-s)}f(s)ds=\int_0^\infty e^{qs}f(s)ds\int_s^\infty e^{-(\lambda+q)t}dt$$

$$=\frac{1}{\lambda+q}\int_0^\infty e^{-\lambda s}f(s)ds=\frac{\phi(\lambda)}{\lambda+q}.$$

在上述计算过程中以$|f(s)|$代$f(s)$，由$\int_0^\infty e^{-\lambda s}|f(s)|ds<\infty$可知，在计算中交换积分号是可行的，且$\int_0^\infty e^{-\lambda t}|g(t)|dt<\infty$. □

现在回到转移概率函数上来．由于$p_{ij}(t)$是$[0,\infty)$上的有界连续函数，它有拉普拉斯变换，记为

$$\psi_{ij}(\lambda)=\int_0^\infty e^{-\lambda t}p_{ij}(t)dt,\quad \lambda>0,\quad i,j\in E.$$

它们也称为转移概率函数的预解式．

定理 2.3 对一切$i,j\in E$及$\lambda,\mu>0$，

(1) $\psi_{ij}(\lambda)\geqslant 0;$ (9)

(2) $\lambda\sum_j\psi_{ij}(\lambda)=1;$ (10)

(3) $\psi_{ij}(\lambda)-\psi_{ij}(\mu)+(\lambda-\mu)\sum_k\psi_{ik}(\lambda)\psi_{kj}(\mu)=0;$ (11)

(4) $\lim\limits_{\lambda\to\infty}\lambda\psi_{ij}(\lambda)=\delta_{ij};$ (12)

(5) $\lim\limits_{\lambda\to\infty}\lambda\{\lambda\psi_{ij}(\lambda)-\delta_{ij}\}=q_{ij}.$ (13)

证 (1) 是明显的．(2) 由条件$\sum_j p_{ij}(t)=1$即得．(3)可由科尔莫戈罗夫－查普曼方程用如下方法推得：

$$\int_0^\infty\int_0^\infty e^{-(\lambda i+\mu s)}p_{ij}(t+s)dtds=\int_0^\infty\int_0^\infty e^{-(\lambda t+\mu s)}\sum_k p_{ik}(t)p_{kj}(s)dtds$$

$$=\sum_k\psi_{ik}(\lambda)\psi_{kj}(\mu).\tag{14}$$

现在不妨设$\lambda>\mu$．注意到

$$\int_0^\infty e^{-\mu s}p_{ij}(t+s)ds=\int_t^\infty e^{-\mu(u-t)}p_{ij}(u)du$$

$$=e^{\mu t}\left\{\int_0^\infty e^{-\mu u}p_{ij}(u)du-\int_0^t e^{-\mu u}p_{ij}(u)du\right\}$$

$$=e^{\mu t}\psi_{ij}(\mu)-\int_0^t e^{\mu(t-u)}p_{ij}(u)du,$$

我们有

$$\int_0^\infty\int_0^\infty e^{-(\lambda t+\mu s)}p_{ij}(t+s)dtds=\int_0^\infty e^{-\lambda t}dt\int_0^\infty e^{-\mu s}p_{ij}(t+s)ds$$

$$
\begin{aligned}
&= \psi_{ij}(\mu)\int_0^\infty e^{-(\lambda-\mu)t}dt - \int_0^\infty e^{-\mu u}p_{ij}(u)du\int_u^\infty e^{-(\lambda-\mu)t}dt \\
&= \frac{\psi_{ij}(\mu)}{\lambda-\mu} - \int_0^\infty e^{-\lambda u}p_{ij}(u)\frac{du}{\lambda-\mu} \\
&= -\frac{1}{\lambda-\mu}\{\psi_{ij}(\lambda)-\psi_{ij}(\mu)\}.
\end{aligned} \tag{15}
$$

易见,若 $\mu>\lambda$,我们可得到同样的结果.由(14)及(15)式即得(11)式.

(4) 可由连续性条件推得.事实上,

$$
\lambda\psi_{ij}(\lambda)=\lambda\int_0^\infty e^{-\lambda t}p_{ij}(t)dt=\int_0^\infty e^{-u}p_{ij}(u/\lambda)du,
$$

$$
\lim_{\lambda\to\infty}p_{ij}(u/\lambda)=\delta_{ij},\qquad \int_0^\infty e^{-u}du=1,
$$

由控制收敛定理即得(12)式.

最后证明(5),我们有

$$
\begin{aligned}
\lambda\{\lambda\psi_{ij}(\lambda)-\delta_{ij}\} &= \int_0^\infty \lambda^2 e^{-\lambda t}[p_{ij}(t)-\delta_{ij}]dt \\
&= \int_0^\infty \lambda e^{-u}[p_{ij}(u/\lambda)-\delta_{ij}]du \\
&= \int_0^\infty u e^{-u}\frac{p_{ij}(u/\lambda)-\delta_{ij}}{u/\lambda}du,
\end{aligned}
$$

$$
\lim_{\lambda\to\infty}\frac{p_{ij}(u/\lambda)-\delta_{ij}}{u/\lambda}=q_{ij},\qquad \left|\frac{p_{ij}(u/\lambda)-\delta_{ij}}{u/\lambda}\right|\leqslant q_i,
$$

及 $\int_0^\infty ue^{-u}du=1$,仍由控制收敛定理得到(13)式. □

把预解式写成矩阵 $\boldsymbol{\Psi}(\lambda)=(\psi_{ij}(\lambda))$,定理 2.3 中的预解式的性质也有简洁的形式.如(11)式为

$$
\boldsymbol{\Psi}(\lambda)-\boldsymbol{\Psi}(\mu)+(\lambda-\mu)\boldsymbol{\Psi}(\lambda)\boldsymbol{\Psi}(\mu)=\boldsymbol{0},
$$

通常称为预解方程.(12)及(13)式分别为

$$
\lim_{\lambda\to\infty}\lambda\boldsymbol{\Psi}(\lambda)=\boldsymbol{I},\quad \lim_{\lambda\to\infty}\lambda\{\lambda\boldsymbol{\Psi}(\lambda)-\boldsymbol{I}\}=\boldsymbol{Q}.
$$

若给出函数 $\boldsymbol{\Psi}(\lambda)=(\psi_{ij}(\lambda))$,$\lambda>0$,满足(9)、(10)、(11)及(12),那么可以证明,这时存在标准转移概率函数,以 $\boldsymbol{\Psi}(\lambda)$为它的预解式.今后我们不需要用到这个结论,所以不给出证明.

我们指出,如果 $p_{ij}(t)$是标准广转移概率函数,它的预解式 $\psi_{ij}(\lambda)$同样满足定理 2.3,只是其中的(10)式改为

$$
\lambda\sum_j\psi_{ij}\lambda\leqslant 1.
$$

对向后方程(6)取拉普拉斯变换,利用引理 2.3,得到预解式满足的向后方程(线性代数方程组)

$$\psi_{ij}(\lambda)=\sum_{k\neq i}\frac{q_{ik}}{\lambda+q_i}\psi_{kj}(\lambda)+\frac{\delta_{ij}}{\lambda+q_i},\quad \lambda>0,\quad i,j\in E, \tag{16}$$

由引理 2.2,(16)与(6)是等价的.

同样地,对向前方程(7)取拉普拉斯变换,得到预解式满足的向前方程(线性代数方程组)

$$\psi_{ij}(\lambda)=\sum_{k\neq j}\psi_{ik}(\lambda)\frac{q_{ik}}{\lambda+q_j}+\frac{\delta_{ij}}{\lambda+q_j},\quad \lambda>0,\quad i,j\in E. \tag{17}$$

也能证明(17)与(7)是等价的.

习　　题

2-1. 设马尔可夫链有两个状态,且它们不全为吸收状态,则转移概率函数矩阵为:

$$\mathbf{P}(t)=\frac{1}{\lambda+\mu}\begin{bmatrix}\mu+\lambda e^{-(\lambda+\mu)t} & \lambda-\lambda e^{-(\lambda+\mu)t}\\ \mu-\mu e^{-(\lambda+\mu)t} & \lambda+\mu e^{-(\lambda+\mu)t}\end{bmatrix},$$
$$\lambda\geqslant 0,\mu\geqslant 0,\lambda+\mu>0.$$

2-2. 设 $\boldsymbol{A}$ 为随机矩阵:

$$\boldsymbol{A}=\begin{pmatrix}\alpha & 1-\alpha\\ 1-\beta & \beta\end{pmatrix},\quad 0\leqslant\alpha,\beta\leqslant 1,$$

则存在转移概率函数矩阵$\{\mathbf{P}(t),t\geqslant 0\}$及 $t_0>0$ 使得 $\boldsymbol{A}=\mathbf{P}(t_0)$的充要条件是 $\alpha+\beta>1$.

2-3. 设马尔可夫链有两个状态,且对一切 $t\geqslant 0$

$$\mathsf{P}(X_t=0)=\mathsf{P}(X_t=1)=1/2,$$

则其转移概率矩阵为

$$\mathbf{P}(t)=\frac{1}{2}\begin{bmatrix}1+e^{-\lambda t} & 1-e^{-\lambda t}\\ 1-e^{-\lambda t} & 1+e^{-\lambda t}\end{bmatrix},\quad \lambda\geqslant 0.$$

2-4. 设密度矩阵为

$$\boldsymbol{Q}=\begin{bmatrix}-1 & 1/2 & 1/2\\ 1/2 & -1 & 1/2\\ 1/2 & 1/2 & -1\end{bmatrix},$$

求转移概率函数$\{\mathbf{P}(t),t\geqslant 0\}$.

2-5. 设状态空间为$\{0,1,\cdots,a\}$的马尔可夫链的密度矩阵为

$$\boldsymbol{Q}=\begin{bmatrix}-\lambda & \lambda & 0 & 0 & \cdots & 0\\ 0 & -\lambda & \lambda & 0 & \cdots & 0\\ \cdots & \cdots & \cdots & \ddots & \cdots & \cdots\\ 0 & \cdots & \cdots & \cdots & -\lambda & \lambda\\ 0 & \cdots & \cdots & \cdots & 0 & 0\end{bmatrix},$$

求转移概率函数$\{\mathbf{P}(t), t\geqslant 0\}$.

2-6. 设状态空间为$\{0,1,\cdots,a\}$的生灭过程的出生率与死亡率为:

$$\lambda_i=(a-i)\lambda,\quad \mu_i=i\mu,\quad 0\leqslant i\leqslant a,\quad \lambda,\mu>0.$$

这个生灭过程是连续时间的埃伦弗斯特扩散模型(参见第二章的习题 5-2.). 设 $X_0=0$. 证明: 对 $t>0$

$$\mathsf{P}(X_t=i)=\mathrm{C}_a^i q_t^i q_t^{a-i},$$

其中

$$p_t=\frac{\lambda}{\lambda+\mu}-\frac{\lambda}{\lambda+\mu}\mathrm{e}^{-(\lambda+\mu)t},\quad q_t=\frac{\mu}{\lambda+\mu}+\frac{\lambda}{\lambda+\mu}\mathrm{e}^{-(\lambda+\mu)t}.$$

§3.3 最小解与规则链

在这一节中,我们先直接讨论科尔莫戈罗夫方程的解. 所以我们讨论的出发点就是密度矩阵 $Q=(q_{ij})$. 实际上,我们就从一个 Q 矩阵(q_{ij})出发进行讨论.

对任意的 $i,j\in E$ 及 $t\geqslant 0$,定义

$$\begin{aligned}
f_{ij}^{(0)}(t)&=\delta_{ij}\mathrm{e}^{-q_i t},\\
f_{ij}^{(n+1)}(t)&=\sum_{k\neq i}\int_0^t \mathrm{e}^{-q_i(t-s)}q_{ik}f_{kj}^{(n)}(s)\mathrm{d}s,\quad n\geqslant 0,\\
f_{ij}(t)&=\sum_{n=0}^{\infty}f_{ij}^{(n)}.
\end{aligned}$$

这实际上是用逐步逼近法解积分方程. 记 $f_{ij}^{(n)}(t)$的拉普拉斯变换为 $\phi_{ij}^{(n)}(\lambda)$,则

$$\begin{aligned}
\phi_{ij}^{(0)}(\lambda)&=\frac{\delta_{ij}}{\lambda+q_i},\\
\phi_{ij}^{(n+1)}(\lambda)&=\sum_{k\neq i}\frac{q_{ij}}{\lambda+q_i}\phi_{kj}^{(n)}(\lambda),\quad n\geqslant 0,\\
\phi_{ij}(\lambda)&=\sum_{n=0}^{\infty}\phi_{ij}^{(n)}(\lambda).
\end{aligned}$$

$\phi_{ij}(\lambda)$为 $f_{ij}(t)$的拉普拉斯变换. 另一方面,由 $\phi_{ij}(\lambda)$的定义可知,它是线性方程组

$$Z_{ij}=\sum_{k\neq i}\frac{q_{ik}}{\lambda+q_i}Z_{kj}+\frac{\delta_{ij}}{\lambda+q_i},\qquad i,j\in E,\quad \lambda>0 \tag{1}$$

的最小非负解. 因此可以期望$\{f_{ij}(t), i,j\in E\}$为向后方程的一个最小解. 自然,我们要的解应是转移概率函数才有意义,然而$\{f_{ij}(t), i,j\in E\}$一般只是广转移概率函数(这也是必须引进广转移概率函数这个概念的一个原因). 这就是下述定理的内容.

定理 3.1 $\{f_{ij}(t), i,j\in E\}$为标准广转移概率函数,且满足向后方程:

$$f_{ij}(t)=\delta_{ij}\mathrm{e}^{-q_i t}+\sum_{k\neq i}\int_0^t \mathrm{e}^{-q_i(i-s)}q_{ik}f_{kj}(s)\mathrm{d}s,\qquad i,j\in E,\quad t\geqslant 0.$$

设$\{g_{ij}(t), i,j\in E\}$也为广转移概率函数，且

$$g_{ij}(t)=\delta_{ij}e^{-q_it}+\sum_{k\neq i}\int_0^t e^{-q_i(t-s)}q_{ik}g_{kj}(s)ds,\quad i,j\in E,\quad t\geqslant 0,$$

则对一切 $i,j\in E$ 及 $t\geqslant 0$，

$$g_{ij}(t)\geqslant f_{ij}(t).$$

证 $f_{ij}(t)\geqslant 0$ 是显然的.我们要证明对一切 $i\in E$，

$$\sum_j f_{ij}(t)\leqslant 1,\qquad t\geqslant 0. \tag{2}$$

记 $\tilde{f}_{ij}^{(n)}(t)=\sum_{v=0}^n f_{ij}^{(v)}(t)$，则 $\tilde{f}_{ij}^{(n)}(t)\uparrow f_{ij}(t)$.为了得到(2)式，只要证明对一切 $n\geqslant 0$

$$\sum_j \tilde{f}_{ij}^{(n)}(t)\leqslant 1,\qquad t\geqslant 0. \tag{3}$$

现在用归纳法证明(3)式. $n=0$ 时

$$\sum_j \tilde{f}_{ij}^{(0)}(t)=\sum_j f_{ij}^{(0)}(t)=e^{-q_it}\leqslant 1,\qquad t\geqslant 0.$$

设(3)式对 n 成立，则

$$\begin{aligned}\sum_j \tilde{f}_{ij}^{(n+1)}(t)&=e^{-q_it}+\sum_{k\neq i}\int_0^t e^{-q_i(t-s)}q_{ik}\Big(\sum_j \tilde{f}_{kj}^{(n)}(s)\Big)ds\\&\leqslant e^{-q_it}+\sum_{k\neq i}\int_0^t e^{-q_i(t-s)}q_{ik}ds\\&\leqslant e^{-q_it}+\int_0^t q_ie^{-q_i(t-s)}ds=1.\end{aligned}$$

再证明 $f_{ij}(t)$满足科尔莫戈罗夫－查普曼方程.为此先用归纳法证明，$n\geqslant 0$ 时

$$f_{ij}^{(n)}(t+s)=\sum_{v=0}^n\sum_k f_{ik}^{(v)}(t)f_{kj}^{(n-v)}(s),\qquad t,s\geqslant 0. \tag{4}$$

$n=0$ 时(4)式显然成立.设(4)式对 n 成立，则

$$\begin{aligned}&\sum_{v=0}^{n+1}\sum_k f_{ik}^{(v)}(t)f_{kj}^{(n+1-v)}(s)\\&=\sum_k f_{ik}^{(0)}(t)f_{kj}^{(n+1)}(s)+\sum_{v=1}^{n+1}\sum_k\sum_{l\neq i}\int_0^t e^{-q_iu}q_{il}f_{lk}^{(v-1)}(t-u)f_{kj}^{(n+1-v)}(s)du\\&=e^{-q_it}f_{ij}^{(n+1)}(s)+\sum_{l\neq i}\int_0^t e^{-q_iu}q_{il}\Big(\sum_{v=0}^n\sum_k f_{lk}^{(v)}(t-u)f_{kj}^{(n-v)}(s)\Big)du\\&=\sum_{l\neq i}\int_0^s e^{-q_i(t+u)}q_{il}f_{lj}^{(n)}(s-u)du+\sum_{l\neq i}\int_0^t e^{-q_iu}q_{il}f_{lj}^{(n)}(t-u+s)ds\\&=\sum_{l\neq i}\int_t^{t+s}e^{-q_iu}q_{il}f_{ij}^{(n)}(t+s-u)du+\sum_{l\neq i}\int_0^t e^{-q_iu}q_{il}f_{lj}^{(n)}(t+s-u)ds\\&=\sum_{l\neq i}\int_0^{t+s}e^{-q_iu}q_{il}f_{ij}^{(n)}(t+s-u)du\end{aligned}$$

$$= \sum_{l \neq i} \int_0^{t+s} e^{-q_i(t+s-u)} q_{il} f_{lj}^{(n)}(u) du = f_{ij}^{(n+1)}(t+s),$$

即(4)式对 $n+1$ 也成立.由(4)式可得

$$f_{ij}(t+s) = \sum_{n=0}^{\infty} f_{ij}^{(n)}(t+s) = \sum_{n=0}^{\infty} \sum_{\nu=0}^{n} \sum_k f_{ik}^{(\nu)}(t) f_{kj}^{(n-\nu)}(s)$$

$$= \sum_k \sum_{\nu=0}^{\infty} \sum_{n=\nu}^{\infty} f_{ik}^{(\nu)}(t) f_{kj}^{(n-\nu)}(s) = \sum_k f_{ik}(t) f_{kj}(s).$$

这样就证明了$\{f_{ij}(t), i, j \in E\}$是广转移概率函数.连续性条件是容易验证的:

$$1 \geqslant f_{ii}(t) \geqslant f_{ii}^{(0)}(t) = e^{-q_i t} \to 1, \qquad t \to 0^+,$$

所以 $\lim_{t \to 0^+} f_{ii}(t) = 1$.也不难验证向后方程:

$$f_{ij}(t) = \sum_{n=0}^{\infty} f_{ij}^{(n)}(t) = \delta_{ij} e^{-q_i t} + \sum_{n=0}^{\infty} f_{ij}^{(n+1)}(t)$$

$$= \delta_{ij} e^{-q_i t} + \sum_{k \neq i} \int_0^t e^{-q_i(t-s)} q_{ik} \sum_{n=0}^{\infty} f_{kj}^{(n)}(s) ds$$

$$= \delta_{ij} e^{-q_i t} + \sum_{k \neq i} \int_0^t e^{-q_i(t-s)} q_{ik} f_{kj}(s) ds.$$

最后证明$\{f_{ij}(t), i, j \in E\}$的最小性,即 $g_{ij}(t) \geqslant f_{ij}(t)$.为此需要用归纳法证明.对 $n \geqslant 0$,

$$g_{ij}(t) \geqslant \tilde{f}_{ij}^{(n)}(t), \qquad i, j \in E, \quad t \geqslant 0, \tag{5}$$

然后在(5)式中令 $n \to \infty$ 即行了. $n=0$ 时(5)式显然成立:

$$g_{ij}(t) \geqslant \delta_{ij} e^{-q_i t} = \tilde{f}_{ij}^{(0)}(t).$$

设(5)式对 n 成立,则

$$g_{ij}(t) = \delta_{ij} e^{-q_i t} + \sum_{k \neq i} \int_0^t e^{-q_i(t-s)} q_{ik} g_{kj}(s) ds$$

$$\geqslant f_{ij}^{(0)}(t) + \sum_{k \neq i} \int_0^t e^{-q_i(t-s)} q_{ik} \tilde{f}_{kj}^{(n)}(s) ds$$

$$= \tilde{f}_{ij}^{(n+1)}(t).$$

至此定理全部证毕. □

定理 3.2 $\{f_{ij}(t), i, j \in E\}$满足向前方程

$$f_{ij}(t) = \delta_{ij} e^{-q_j t} + \sum_{k \neq j} \int_0^t f_{ik}(s) q_{kj} e^{-q_j(t-s)} ds.$$

若$\{r_{ij}(t), i, j \in E\}$也为广转移概率函数,且

$$r_{ij}(t) = \delta_{ij} e^{-q_j t} + \sum_{k \neq j} \int_0^t r_{ik}(s) q_{kj} e^{-q_j(t-s)} ds,$$

则对一切 $i, j \in E$ 及 $t \geqslant 0$,

$$r_{ij}(t) \geqslant f_{ij}(t).$$

证 为了得到向前方程的最小解,很自然地也应用逐次逼近法去构造.对任意的 $i,j \in E$ 及 $t \geqslant 0$,定义

$$h_{ij}^{(0)} = \delta_{ij} e^{-q_j t},$$
$$h_{ij}^{(n+1)}(t) = \sum_{k \neq j} \int_0^t h_{ik}^{(n)}(s) q_{kj} e^{-q_j(t-s)} ds, \qquad n \geqslant 0,$$
$$h_{ij}(t) = \sum_{n=0}^{\infty} h_{ij}^{(n)}(t).$$

与定理 3.1 的证明完全一样,可证 $\{h_{ij}(t), i,j \in E\}$ 满足向前方程:

$$h_{ij}(t) = \delta_{ij} e^{-q_j t} + \sum_{k \neq j} \int_0^t h_{ik}(s) q_{kj} e^{-q_j(t-s)} ds,$$

且有最小性:对一切 $i,j \in E$ 及 $t \geqslant 0$, $r_{ij}(t) \geqslant h_{ij}(t)$. 我们只要证明:对一切 $i,j \in E$ 及 $t \geqslant 0$, $h_{ij}(t) = f_{ij}(t)$.

记 $h_{ij}^{(n)}(t)$ 的拉普拉斯变换为 $\psi_{ij}^{(n)}(\lambda)$,则

$$\psi_{ij}^{(0)}(\lambda) = \frac{\delta_{ij}}{\lambda + q_j},$$
$$\psi_{ij}^{(n+1)}(\lambda) = \sum_{k \neq j} \psi_{ik}^{(n)}(\lambda) \frac{q_{kj}}{\lambda + q_j}, \quad n \geqslant 0,$$
$$\psi_{ij}(\lambda) = \sum_{n=0}^{\infty} \psi_{ij}^{(n)}(\lambda).$$

$\psi_{ij}(\lambda)$ 是 $h_{ij}(t)$ 的拉普拉斯变换,它是线性方程组

$$z_{ij} = \sum_{k \neq j} z_{ik} \frac{q_{kj}}{\lambda + q_j} + \frac{\delta_{ij}}{\lambda + q_j}, \quad i,j \in E, \quad \lambda > 0 \tag{6}$$

的最小非负解.

我们只要证明 $\psi_{ij}(\lambda) = \varphi_{ij}(\lambda), i,j \in E, \lambda > 0$,即方程组(1)与(6)有相同的最小非负解.为方便起见,我们用矩阵形式.记

$$\boldsymbol{\Phi} = (\phi_{ij}(\lambda)), \quad \boldsymbol{\Psi} = (\psi_{ij}(\lambda)), \quad \boldsymbol{Z} = (z_{ij}),$$
$$\boldsymbol{C} = \left((1 - \delta_{ij}) \frac{q_{ij}}{\lambda + q_i}\right), \quad \boldsymbol{D} = \left((1 - \delta_{ij}) \frac{q_{ij}}{\lambda + q_j}\right), \quad \boldsymbol{B} = \left(\frac{\delta_{ij}}{\lambda + q_i}\right) = \left(\frac{\delta_{ij}}{\lambda + q_j}\right)$$

(取 $\lambda > 0$ 为固定的,不再标出),则 $\boldsymbol{\Phi}$ 是

$$\boldsymbol{Z} = \boldsymbol{CZ} + \boldsymbol{B}$$

的最小非负解,因此 $\boldsymbol{\Phi} = \sum_{n=0}^{\infty} \boldsymbol{C}^n \boldsymbol{B}$; $\boldsymbol{\Psi}$ 是

$$\boldsymbol{Z} = \boldsymbol{ZD} + \boldsymbol{B}$$

的最小非负解,因此 $\boldsymbol{\Psi} = \sum_{n=0}^{\infty} \boldsymbol{BD}^n$. 容易直接验证

$$CB = \left((1-\delta_{ij}) \frac{q_{ij}}{(\lambda + q_i)(\lambda + q_j)} \right) = BD,$$

从而对一切 $n \geqslant 1$,

$$C^n B = C^{n-1} BD = \cdots = BD^n,$$

由此即知 $\boldsymbol{\Phi} = \boldsymbol{\Psi}$. □

根据定理3.1及3.2,我们称$\{f_{ij}(t), i, j \in E\}$为科尔莫戈罗夫方程的最小解,它是满足向后方程的最小 Q 过程,也是满足向前方程的最小 Q 过程.很自然,我们感兴趣的是最小解确实是转移概率函数的情形.

定义 Q 矩阵(q_{ij})称为规则的,若对一切 $i \in E$ 及 $t \geqslant 0$,

$$\sum_j f_{ij}(t) = 1.$$

若一个马尔可夫链的密度矩阵是规则的,则称这个马尔可夫链是规则的.

定理 3.3 Q 矩阵(q_{ij})为规则的充要条件是:

(1) (q_{ij})是保守的;

(2) 对任意的 $\lambda > 0$,方程组

$$\lambda z_i = \sum_k q_{ik} z_k, \quad 0 \leqslant z_i \leqslant 1, \quad i \in E \tag{7}$$

的解全为零.

证 规则性即对一切 $i \in E$ 及 $t \geqslant 0$, $\sum_j f_{ij}(t) = 1$,这等价于对一切 $i \in E$ 及 $\lambda > 0$, $\lambda \sum_j \phi_{ij}(\lambda) = 1$.我们已经知道 $\sum_j f_{ij}(t) \leqslant 1$,即 $\lambda \sum_j \phi_{ij}(\lambda) \leqslant 1$.令

$$\lambda \sum_j \phi_{ij}(\lambda) = 1 - \psi_i(\lambda), \quad i \in E, \quad \lambda > 0,$$

则 $0 \leqslant \psi_i(\lambda) \leqslant 1, i \in E, \lambda > 0$,且规则性等价于对一切 $i \in E$ 及 $\lambda > 0$, $\psi_i(\lambda) = 0$.

记 $\rho_i^{(n)}(\lambda) = \lambda \sum_j \phi_{ij}^{(n)}(\lambda)$,则

$$\rho_i^{(0)} = \frac{\lambda}{\lambda + q_i},$$

$$\rho_i^{(n+1)} = \sum_{k \neq i} \frac{q_{ik}}{\lambda + q_i} \rho_k^{(n)}(\lambda), \quad n \geqslant 0,$$

$$\sum_{n=0}^{\infty} \rho_i^{(n)}(\lambda) = \lambda \sum_j \phi_{ij}(\lambda) = 1 - \psi_i(\lambda).$$

因此$\{1 - \psi_i(\lambda), i \in E\}$是线性方程组

$$z_i = \sum_{k \neq i} \frac{q_{ik}}{\lambda + q_i} z_k + \frac{\lambda}{\lambda + q_i}, \qquad i \in E$$

的最小非负解,从而$\{\psi_i(\lambda), i \in E\}$是线性方程组

$$z_i = \sum_{k \neq i} \frac{q_{ik}}{\lambda + q_i} z_k - \sum_k \frac{q_{ik}}{\lambda + q_i}, \quad 0 \leqslant z_i \leqslant 1, \quad i \in E \tag{8}$$

的最大解.

先证充分性.由保守性,方程组(7)与方程组(8)一致.方程组(7)的解全为零,自然最大解也为零:$\psi_i(\lambda)=0, i\in E, \lambda>0$,从而密度矩阵是规则的.

再证必要性.$\{\psi_i(\lambda), i\in E\}$是方程组(8)的最大解,但$\psi_i(\lambda)=0, i\in E$,因此方程组(8)必须是齐次的:

$$\sum_k q_{ik}=0,\qquad i\in E,$$

即(q_{ij})是保守的.现在方程组与(7)又一致了.既然方程组(8)的最大解为零,方程组(7)的解也全为零了. □

推论 1 设保守密度 Q 矩阵(q_{ij})为有界的:$q=\sup_i q_i<\infty$,则 Q 是规则的.

证 设$\{u_i, i\in E\}$是方程组(7)的解.记 $u=\sup_i u_i$.对任意的 $\varepsilon>0$,取 $i\in E$,使 $u_i\geqslant u-\varepsilon$,则对 $\lambda>0$,

$$(\lambda+q_i)(u-\varepsilon)\leqslant(\lambda+q_i)u_i=\sum_{k\neq i}q_{ik}u_k\leqslant q_i u,$$
$$\lambda(u-\varepsilon)\leqslant q_i\varepsilon\leqslant q\varepsilon,$$

令 $\varepsilon\to 0$,得 $\lambda u\leqslant 0$,因此 $u=0, u_i=0, i\in E$. □

推论 2 设保守链的跳跃链的每个状态都是常返的,则链是规则的.

证 设$\{u_i, i\in E\}$为方程组(7)的解.若 $q_i=0$,则 $\lambda u_i=0, u_i=0$.记 $G=\{i: q_i>0\}$.由于跳跃链的状态都是常返的,G 与$E\setminus G$ 都是闭集,方程组(7)可归结为

$$(\lambda+q_i)u_i=\sum_{k\neq i, k\in G}q_{ik}u_k,\qquad i\in G.$$

因此不妨设 $G=E$,即对一切 $i\in E, q_i>0$.这时

$$\sum_k r_{ik}u_k=\frac{\lambda+q_i}{q_i}u_i,\qquad i\in E,$$

$n\geqslant 1$ 时,

$$\sum_k r_{ik}^{(n+1)}u_k=\sum_k\sum_l r_{il}^{(n)}r_{lk}u_k=\sum_l r_{il}^{(n)}\sum_k r_{lk}u_k$$
$$=\sum_l r_{il}^{(n)}u_l\frac{\lambda+q_l}{q_l}=\sum_l r_{il}^{(n)}u_l+\lambda\sum_l r_{il}^{(n)}\frac{u_l}{q_l},$$
$$\sum_k r_{ik}^{(n+1)}u_k-\sum_k r_{ik}^{(n)}u_k\geqslant\lambda r_{ii}^{(n)}\frac{u_i}{q_i},\qquad i\in E,$$

再由于 $\sum_k r_{ik}u_k\geqslant\lambda u_i/q_i, i\in E$,可得 $n\geqslant 1$ 时,

$$1\geqslant\sum_k r_{ik}^{(n+1)}u_k\geqslant\lambda\sum_{\upsilon=0}^{n}r_{ii}^{(\upsilon)}\frac{u_i}{q_i},\qquad i\in E.$$

但 $\sum_{\upsilon=0}^{n}r_{ii}^{(\upsilon)}\to\infty, n\to\infty$,因此必须 $u_i=0, i\in E$. □

定理 3.4 规则马尔可夫链的转移概率函数满足科尔莫戈罗夫向前与向后方程,且是方程

的唯一解.

证 由前一定理,规则链是保守的,它的转移概率函数$\{p_{ij}(t),i,j\in E\}$满足向后方程,因此由最小解的最小性,

$$p_{ij}(t)\geqslant f_{ij}(t),\quad i,j\in E,\quad t\geqslant 0.$$

但$1=\sum_j p_{ij}(t)=\sum_j f_{ij}(t)=1$,所以必须对一切$i,j\in E$及$t\geqslant 0$

$$p_{ij}(t)=f_{ij}(t),$$

即规则链的转移概率函数就是最小解.

用完全相同的证明可知,若$\{g_{ij}(t),i,j\in E\}$为标准广转移概率函数,且满足向后方程(或向前方程),则也必须有$g_{ij}(t)=f_{ij}(t),i,j\in E,t\geqslant 0$.所以,规则链的转移概率函数是向后方程(向前方程)的唯一解. □

前面的讨论表明,规则马尔可夫链具有良好的数学性质,然而规则性是通过最小解来定义的,最小解又是用数学中常用的逐次逼近法定义的.那么最小解是否纯粹是数学分析的产物呢?实际上并非如此,最小解与规则性都有清晰的概率意义.

首先,我们指出

$$f_{ij}^{(n)}(t)=\mathsf{P}(X_t=j,\text{且}(0,t)\text{中恰发生 } n \text{ 次状态转移}\mid X_0=i).\tag{9}$$

我们用归纳法来论证它.$n=0$时(9)式是显然的,因为等式的右边即$\delta_{ij}\mathsf{P}(\tau>t\mid X_0=i)=\delta_{ij}\mathrm{e}^{-q_it}=f_{ij}^{(0)}$,这里$\tau$是停留在初始状态上的时间.设(9)式对$n$成立,我们来说明它对$n+1$也成立.若$X_t=j$,且$(0,t)$中发生了$n+1$次状态转移,我们考虑在初始状态$i$上停留的时间$\tau$的取值.$\tau\in(t-s,t-s+\mathrm{d}s)$的概率为$q_i\mathrm{e}^{-q_i(t-s)}\mathrm{d}s$,在时刻$\tau$发生第一次状态转移,从状态$i$转移到状态$k(\neq i)$的概率为$q_{ik}/q_i$,然后从状态$k$经过时间$s$到达状态$j$,但中间发生$n$次状态转移,概率即为$f_{kj}^{(n)}(s)$.这样,

$$\begin{aligned}&\mathsf{P}(X_t=j,\text{且}(0,t)\text{中恰发生 } n+1 \text{ 次状态转移}\mid X_0=i)\\&=\sum_{k\neq i}\int_0^t q_i\mathrm{e}^{-q_i(t-s)}\frac{q_{ik}}{q_i}f_{kj}^{(n)}(s)\mathrm{d}s=f_{ij}^{(n+1)}(t).\end{aligned}$$

由(9)式即得

$$f_{ij}(t)=\mathsf{P}(X_t=j,\text{且}(0,t)\text{中至多发生有限次状态转移}\mid X_0=i),$$
$$\sum_j f_{ij}(t)=\mathsf{P}((0,t)\text{中至多发生有限次状态转移}\mid X_0=i).$$

这样,规则链的定义$\sum_j f_{ij}(t)=1,i\in E$的概率意义就是,马尔可夫链在任意有限区间内至多只发生有限次状态转移.定义

$$\tau_0=0,$$
$$\tau_{n+1}=\inf\{t>\tau_n:X_t\neq X_{\tau_n}\},\quad n\geqslant 0,$$

$\tau_n,n\geqslant 1$就是发生第n次状态转移的时刻,或者第n次跳跃的时刻.由于在有限区间内至多只

发生有限次跳跃,我们有 $\tau_n \uparrow \infty$. 在 τ_n 与 τ_{n+1}之间状态保持不变:在(τ_0,τ_1)中停留于初始状态 X_0,在(τ_1,τ_2)中停留于状态 X_{τ_i},…….所以,规则链可以如下地构造:

$$X_t = \xi_n (= X_{\tau_n}), \qquad t \in [\tau_n, \tau_{n+1}), \quad n \geqslant 0.$$

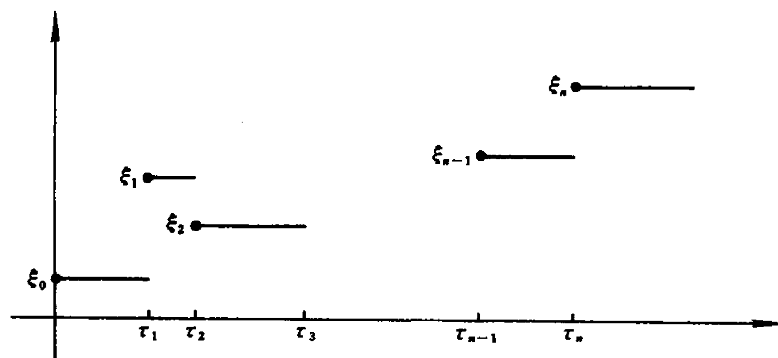


图 2

换句话说,规则马尔可夫链的样本函数是阶梯函数,且在任意有限区间内至多只有有限个不连续点(跳跃点),它由跳跃时刻$\{\tau_n, n \geqslant 1\}$和在跳跃时刻的状态$\{\xi_n, n \geqslant 1\}$及初始状态完全决定.状态的跳跃由跳跃链决定,跳跃时刻由相继的停留时间确定.在已知跳跃链的状态的条件下,在这些状态停留的时间相互独立,且服从指数分布.因此,在没有吸收状态时,它们的有限维分布为:$0 < t_1 < t_2 < \cdots < t_n$ 时

$$\begin{aligned}
&\mathrm{P}(\xi_0 = i_0, \tau_1 \in (t_1, t_1 + \mathrm{d}t_1), \xi_1 = i_1, \cdots, \tau_n \in (t_n, t_n + \mathrm{d}t_n), \xi_n = i_n) \\
&\quad = p_{i_0} q_{i_0} \mathrm{e}^{-q_0 t_1} \mathrm{d}t_1 \frac{q_{i_0 i_1}}{q_{i_0}} \cdots q_{i_{n-1}} \mathrm{e}^{-q_{in-1}(t_n - t_{n-1})} \mathrm{d}t_n \frac{q_{i_{n-1} i_n}}{q_{i_{n-1}}} \\
&\quad = p_{i_0} q_{i_0 i_1} \cdots q_{i_{n-1} i_n} \exp\{-[q_{i_0} t_1 + q_{i_1}(t_2 - t_1) + \cdots + q_{i_{n-1}}(t_n - t_{n-1})]\} \mathrm{d}t_1 \cdots \mathrm{d}t_n.
\end{aligned}$$

若存在吸收状态,则 ξ_n 为吸收状态时,$\tau_{n+1} = \infty$,ξ_{n+1}也不必定义了.这时有限维分布也不难写出,只是形式上更复杂些吧了.综上所述,对规则链我们已经完全掌握了它的结构.由于有这样简单的结构,使进一步分析状态的性质及状态转移的规律大为简化了.

定理 3.5 设生灭过程的密度矩阵为

$$\boldsymbol{Q} = \begin{pmatrix}
-\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\
\mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots \\
0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix},$$

$$\lambda_i \geqslant 0, \quad i \geqslant 0; \quad \mu_i \geqslant 0, \quad i \geqslant 1.$$

(1) 若有无穷多个 n,使 $\lambda_n = 0$,则生灭过程是规则的;

(2) 若至多有有限个 n,使 $\lambda_n = 0$,令

$$a=\begin{cases}\max\{n:\lambda_n=0\}+1, & 若\{n:\lambda_n=0\}非空集,\\ 0, & 若\{n:\lambda_n=0\}为空集,\end{cases}$$

则生灭过程为规则的充要条件是

$$\sum_{n=a+1}^{\infty}\left(\frac{1}{\lambda_n}+\frac{\mu_n}{\lambda_n\lambda_{n-1}}+\cdots+\frac{\mu_n\cdots\mu_{a+1}}{\lambda_n\cdots\lambda_a}\right)=\infty .$$

证 对生灭过程,方程组(7)形为

$$\begin{cases}(\lambda+\lambda_0)z_0=\lambda_0 z_1,\\ (\lambda+\lambda_i+\mu_i)z_i=\mu_i z_{i-1}+\lambda_i z_{i+1}, & i\geqslant 1,\\ 0\leqslant z_i\leqslant 1, & i\geqslant 0.\end{cases}\tag{10}$$

若 $\lambda_b=0$,则前 $b+1$ 个方程构成下列方程组:

$$\begin{cases}(\lambda+\lambda_0)z_0=\lambda_0 z_1,\\ (\lambda+\lambda_i+\mu_i)z_i=\mu_i z_{i-1}+\lambda_i z_{i+1}, & 1\leqslant i\leqslant b-1,\\ (\lambda+\mu_b)z_b=\mu_b z_{b-1},\\ 0\leqslant z_i\leqslant 1, & 0\leqslant i\leqslant b.\end{cases}\tag{11}$$

如果考虑 $b+1$ 维的保守 Q 矩阵

$$Q_b=\begin{pmatrix}-\lambda_0 & \lambda_0 & 0 & \cdots & \cdots & 0\\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & \cdots & \cdots & 0\\ \vdots & & & \ddots & & \vdots\\ 0 & \cdots & \cdots & \mu_{b-1} & -(\lambda_{b-1}+\mu_{b-1}) & \lambda_{b-1}\\ 0 & \cdots & \cdots & 0 & \mu_b & -\mu_b\end{pmatrix},$$

Q_b 自然是规则的.对应于 Q_b 的方程组(7)即方程组(11),因此(11)的解全为零.由此可知,若 $\{u_i,i\in E\}$ 为方程组(10)的解,只要 $\lambda_b=0$,就有 $u_0=\cdots=u_b=0$.这样,定理的第一个结论(1)就得证了.

再证第二个结论(2).为方便起见,规定 $\mu_0=0$,并引进记号:

$$f_n=\frac{1}{\lambda_n}>0,\quad g_n=\frac{\mu_n}{\lambda_n},\quad n\geqslant a,$$

$$m_n=\begin{cases}f_a, & n=a,\\ f_n+g_nf_{n-1}+\cdots+g_n\cdots g_{a+1}f_a, & n\geqslant a+1.\end{cases}$$

先证充分性.设 $\sum_{n=a}^{\infty}m_n=\infty$.设 $\{u_i,i\in E\}$ 为方程组(10)的解,则 $u_0=\cdots=u_{a-1}=0$,

$$\begin{aligned}u_{a+1}-u_a&=\lambda f_a u_a+g_a u_a,\\ u_{n+1}-u_n&=\lambda f_n u_n+g_n(u_n-u_{n-1})\\ &=\lambda f_n u_n+g_n[\lambda f_{n-1}u_{n-1}+g_{n-1}(u_{n-1}-u_{n-2})]\\ &=\cdots\end{aligned}$$

$$=\lambda[f_nu_n+g_nf_{n-1}u_{n-1}+\cdots+g_n\cdots g_{a+2}f_{a+1}u_{a+1}]+g_n\cdots g_{a+1}(u_{a+1}-u_a)$$
$$=\lambda[f_nu_n+g_nf_{n-1}u_{n-1}+\cdots+g_n\cdots g_{a+1}f_au_a]+g_n\cdots g_au_a,\quad n\geqslant a+1. \tag{12}$$

易见$\{u_n,n\geqslant 0\}$为单调上升数列(只需用归纳法证明).设 $u_0=\cdots=u_{b-1}=0,u_b>0(b\geqslant a)$.由(12),$n\geqslant b+1$ 时

$$u_{n+1}-u_n\geqslant\lambda g_n\cdots g_{b+1}f_bu_b,$$
$$1\geqslant u_{n+1}-u_b\geqslant\lambda u_bf_b\left(\sum_{v=b+1}^{n}g_v\cdots g_{b+1}\right),$$

从而 $\sum_{n=b+1}^{\infty}g_n\cdots g_{b+1}<\infty$,进而

$$\sum_{n=k}^{\infty}g_n\cdots g_k<\infty,\quad 1\leqslant k\leqslant b.$$

定义

$$m'_n=\begin{cases}f_b, & n=b,\\ f_n+g_nf_{n-1}+\cdots+g_n\cdots g_{b+1}f_b, & n\geqslant b+1,\end{cases}$$

则 $n\geqslant b+1$ 时,

$$m_n-m'_n=\sum_{k=a+1}^{b}g_n\cdots g_kf_{k-1},$$
$$\sum_{n=b+1}^{\infty}(m_n-m'_n)=\sum_{k=a+1}^{b}\sum_{n=b+1}^{\infty}g_n\cdots g_kf_{k-1}<\infty,$$

因此 $\sum_{n=b+1}^{\infty}m'_n=\infty$.仍由(12)式,$n\geqslant b+1$ 时

$$u_{n+1}-u_n\geqslant\lambda m'_nu_b,$$
$$1\geqslant u_{n+1}-u_{b+1}\geqslant\lambda u_b\sum_{v=b+1}^{n}m'_v,$$

从而 $\sum_{n=b+1}^{\infty}m'_n<\infty$,得出矛盾.所以,$\{u_n,n\geqslant a\}$也全为零.

再证必要性.设 $m=\sum_{n=a}^{\infty}m_n<\infty$,自然也有 $g=\sum_{n=a}^{\infty}g_n\cdots g_a<\infty$.定义 $u_0=\cdots=u_{a-1}=0,u_a=e^{-(\lambda m+g)}$,并按(12)式递归地定义$\{u_n,n\geqslant a+1\}$.易见,$0<u_a<1$,且$\{u_n,n\geqslant 0\}$仍是单调上升数列.这些定义的$\{u_i,i\in E\}$满足方程组(10),只差最后一个条件:$0\leqslant u_i\leqslant 1,i\in E$.若我们证明了 $0\leqslant u_i\leqslant 1,n\geqslant a+1$,则方程组(10)有非零解,链不是规则的.下面就证明这一事实.

由(12)式及数列的单调性,$n\geqslant a$ 时,

$$u_{n+1}-u_n\leqslant(\lambda m_n+g_n\cdots g_a)u_n,$$
$$\begin{aligned}u_{n+1}&\leqslant(1+\lambda m_n+g_n\cdots g_a)u_n\\&\leqslant\cdots\\&\leqslant\prod_{v=a}^{n}(1+\lambda m_v+g_v\cdots g_a)u_a\end{aligned}$$

$$\leqslant \exp\left\{\sum_{v=a}^{n}(\lambda m_v + g_v \cdots g_a)\right\} u_a$$

$$= \exp\left\{-\sum_{v=n+1}^{\infty}(\lambda m_v + g_v \cdots g_a)\right\} < 1$$

(这里利用了不等式 $1+x \leqslant e^x, x \geqslant 0$). 至此定理全部证毕. □

推论 1 纯灭过程是规则的.

推论 2 纯生过程是规则的充要条件为,对一切 $n \geqslant 0$

$$\sum_{k=n}^{\infty}\frac{1}{\lambda_k} = \infty \qquad \left(\text{规定}\frac{1}{0} = \infty\right).$$

例 3.1 线性纯生过程 这时 $\lambda_n = n\lambda, n \geqslant 0$. 显然, $\sum_{n=1}^{\infty} 1/\lambda_n = \infty$,因此线性纯生过程是规则的. 为了求得转移概率函数,我们解向前方程.

首先指出, $j < i$ 时, $p_{ij}(t) \equiv 0$(这符合于只生不灭的直观意义). 事实上, $i > 0$ 时先有

$$p'_{i0}(t) = -\lambda_0 p_{i0}(t), \qquad p_{i0}(0) = 0,$$

因此, $p_{i0}(t) \equiv 0$. 这样

$$p'_{i1}(t) = \lambda_0 p_{i0}(t) - \lambda_1 p_{i1}(t) = -\lambda_1 p_{i1}(t),$$

若 $i > 1$,则 $p_{i1}(0) = 0$,也有 $p_{i1}(t) \equiv 0$. 逐步推下去,直至得到 $p_{i,i-1}(t) \equiv 0$.

现在,向前方程归结为

$$\begin{cases} p'_{ii}(t) = -\lambda_i p_{ii}(t), \\ p'_{ij}(t) = \lambda_{j-1} p_{i,j-1}(t) - \lambda_j p_{ij}(t), \quad j \geqslant i+1. \end{cases}$$

注意到

$$[e^{\lambda_j t} p_{ij}(t)]' = e^{\lambda_j t} p'_{ij}(t) + \lambda_j e^{\lambda_j t} p_{ij}(t) = \lambda_{j-1} p_{i,j-1}(t) e^{\lambda_j t},$$

我们得到递推公式

$$\begin{cases} p_{ii}(t) = e^{-\lambda_i t}, \\ p_{ij}(t) = \lambda_{j-1} e^{-\lambda_j t} \int_0^t e^{\lambda_j s} p_{i,j-1}(s) \mathrm{d}s, j \geqslant i+1. \end{cases} \tag{13}$$

(13)式对一般的纯生过程都成立.

对于线性纯生过程,我们有

$$\begin{cases} p_{00}(t) \equiv 1, \quad p_{0j}(t) \equiv 0, & j > 1, \\ p_{ij}(t) = C_{j-1}^{j-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{j-i}, & i \geqslant 1, j \geqslant i, \\ p_{ij}(t) \equiv 0, & i \geqslant 1, j < i. \end{cases} \tag{14}$$

由(13)式及归纳法就能证明(14)式. 事实上, $i \geqslant 1, j \geqslant i$ 时

$$p_{i,j+1}(t) = j\lambda \int_0^t e^{-(t-s)(j+1)\lambda} p_{ij}(s) \mathrm{d}s$$

$$
\begin{aligned}
&= j\lambda \int_0^t \mathrm{e}^{-(t-s)(j+1)\lambda} \mathrm{C}_{j-1}^{j-i} \mathrm{e}^{-i\lambda s}(1-\mathrm{e}^{-\lambda s})^{j-i} \mathrm{d}s \\
&= j\mathrm{C}_{j-1}^{j-i} \mathrm{e}^{-(j+1)\lambda t} \int_0^t \mathrm{e}^{\lambda s}(\mathrm{e}^{\lambda s}-1)^{j-i} \mathrm{d}s \\
&= \frac{j!}{(j-i)!\ (i-1)!} \mathrm{e}^{-(j+1)\lambda t} \frac{(\mathrm{e}^{\lambda t}-1)^{j-i+1}}{j-i+1} \\
&= \mathrm{C}_j^{j+1-i} \mathrm{e}^{-i\lambda t}(1-\mathrm{e}^{-\lambda t})^{j+1-i}.
\end{aligned}
$$

事先不知道(14)式而要通过(13)把它归纳出来还是比较困难的.下面我们利用母函数及解一阶偏微分方程的方法求转移概率函数.不熟悉解一阶偏微分方程的方法的读者可以跳过下面这一段.

令 $\Phi_i(t,z) = \sum_{j=0}^{\infty} p_{ij}(t) z^j, i \geqslant 0, t \geqslant 0, |z| < 1$.线性纯生过程的向前方程为

$$
\begin{cases}
p'_{i0}(t) = 0, \\
p'_{ij}(t) = \lambda(j-1) p_{i,j-1}(t) - \lambda j p_{ij}(t), \qquad j \geqslant 1,
\end{cases}
$$

由此不难得到

$$
\begin{cases}
\dfrac{\partial \Phi_i}{\partial t} = \displaystyle\sum_{j=0}^{\infty} p'_{ij}(t) z^j = \lambda \sum_{j=1}^{\infty} [(j-1) p_{i,j-1}(t) - j p_{ij}(t)] z^j = \lambda z(z-1) \dfrac{\partial \Phi_i}{\partial z}, \\
\Phi_i = (0,z) = z^i.
\end{cases}
$$

由常微分方程

$$
\mathrm{d}t = \frac{\mathrm{d}z}{\lambda z(z-1)} = 0,
$$

求得首次积分

$$
\left(\frac{z}{z-1}\right) \mathrm{e}^{-\lambda t} = C.
$$

因此,一阶偏微分方程的通解为

$$
\Phi_i(t,z) = \psi_i\left(\frac{z}{z-1} \mathrm{e}^{-\lambda t}\right).
$$

由初始条件

$$
\Phi_i(0,z) = \psi_i\left(\frac{z}{z-1}\right) = z^i,
$$

可解得

$$
\psi_i(z) = \left(\frac{z}{z-1}\right)^i,
$$

从而 $i \geqslant 1$ 时

$$
\Phi_i(t,z) = \left[\frac{z\mathrm{e}^{-\lambda t}}{1-(1-\mathrm{e}^{-\lambda t})z}\right]^i = \mathrm{e}^{-i\lambda t} z^i \sum_{n=0}^{\infty} \mathrm{C}_{n+i-1}^{n} (1-\mathrm{e}^{-\lambda t})^n z^n,
$$

这里用到$\left(\sum_{n=0}^{\infty} z^n\right)^i = \sum_{n=0}^{\infty} C_{n+i-1}^n z^n$，$|z|<1$，再比较等式两边幂级数的系数，即可得到(14)式. □

例 3.2 分支过程 现在用连续时间马尔可夫链来描述群体的"繁殖"问题. 以 X_t 表示时刻 t 群体中个体的个数，与离散时间情形不同，"世代"的概念在连续时间情形已不适用了. $p_{ij}(t)$是 i 个个体在时间 t 之后变成 j 个个体的概率，我们假设

$$0<p_{10}(t)<1, \tag{15}$$

不然的话，讨论也要变成平庸的了. 我们仍假设各个个体繁衍后代的状况是相互独立的，而 0 是吸收状态. 因此，记 $\Phi_i(t,z)=\sum_{j=0}^{\infty} p_{ij}(t)z^j$，则应有

$$\Phi_i(t,z)=[\Phi_1(t,z)]^i, \qquad i\geqslant 0. \tag{16}$$

我们称满足条件(16)的连续时间马尔可夫链为**连续时间的分支过程**. 为方便起见，今后记 $G_t(z)=\Phi_1(t,z)$. 这时对 $t,s\geqslant 0$，

$$G_{t+s}(z)=G_t(G_s(z)). \tag{17}$$

事实上，这是科尔莫戈罗夫－查普曼方程的推论：

$$\begin{aligned}
G_{t+s}(z) &= \sum_j p_{1j}(t+s)z^j = \sum_j \sum_k p_{1k}(t)p_{kj}(s)z^j \\
&= \sum_k p_{1k}(t)\sum_j p_{kj}(s)z^j = \sum_k p_{1k}(t)\Phi_k(s,z) \\
&= \sum_k p_{1k}(t)[G_s(z)]^k = G_t(G_s(z)).
\end{aligned}$$

(17)的直观意义也是十分清楚的. 一个个体经过时间 t 变成若干个个体，这些个体再相互独立地经过时间 s 繁殖各自的后裔，它们的总和就是一个个体经过时间 $t+s$ 所繁殖的子孙，因此就有母函数的迭代式.

现在考察分支过程的密度矩阵. $i\geqslant 1$ 时，

$$\begin{aligned}
\sum_{k=0}^{\infty} q_{ik}z^k &= \left.\frac{\partial\Phi_i}{\partial t}\right|_{t=0} = \left.\left(i\Phi_1^{i-1}\frac{\partial\Phi_1}{\partial t}\right)\right|_{t=0} \\
&= iz^{i-1}\sum_{j=0}^{\infty} q_{1j}z^j = \sum_{k=i-1}^{\infty} iq_{1,k-i+1}z^k,
\end{aligned}$$

而 $q_{0j}\equiv 0$；因此对一切 $i,j\in E$，

$$q_{ij}=\begin{cases}0, & j<i-1,\\ iq_{1,j-i+1}, & j\geqslant i-1,\end{cases} \tag{18}$$

特别，$q_i=iq_1$. 由假设(15)，$q_1>0$. 记

$$f(z)=\sum_{j=0}^{\infty} q_{1j}z^j.$$

我们将只讨论保守的分支过程，因此 $f(1)=0$. 这时由于 $f''(z)\geqslant 0$，$f(z)$在(0,1)中保持定号，没有零点，或只有一个零点.

注意到$\left.\frac{\mathrm{d}G_t(z)}{\mathrm{d}t}\right|_{t=0}=f(z)$，由(17)式对 $t\geqslant 0$ 有

$$\frac{\mathrm{d}G_t(z)}{\mathrm{d}t}=f(G_t(z)),$$

$$\int_z^{G_t(z)}\frac{\mathrm{d}x}{f(x)}=t. \tag{19}$$

取 $\delta>0$ 充分小，则 $f(x)$ 在 $(1-\delta,1)$ 中保持定号. 固定 t，当 z 充分接近于 1 时，$z, G_t(z)\in(1-\delta,1)$. 因此

$$\int_{1-\delta}^1\frac{\mathrm{d}x}{|f(x)|}\geqslant\int_z^{G_t(z)}\frac{\mathrm{d}x}{f(x)}=t,$$

令 $t\to\infty$，即有

$$\int_{1-\delta}^1\frac{\mathrm{d}x}{|f(x)|}=\infty. \tag{20}$$

在(19)式中取 $z=0$，由于 $G_t(0)=p_{10}(t)>0$，积分 $\int_0^{G_t(0)}\frac{\mathrm{d}x}{f(x)}$ 收敛，必须有 $q_{10}=f(0)>0$.

一个保守 Q 矩阵若满足条件(18)，且 $q_1>0, q_{10}>0$，就称为分支矩阵. 这样，我们证明了保守分支过程的密度矩阵是分支矩阵，且满足条件(20).

现在，我们证明满足条件(20)的分支矩阵是规则的，且它的最小解也是分支过程，即满足条件(17). 为此只要讨论最小解 $\{f_{ij}(t), i,j\in E\}$ 的性质. 它满足向前方程：

$$f'_{ij}(t)=\sum_{k=0}^{j+1}f_{ik}(t)q_{kj}=\sum_{k=1}^{j+1}kf_{ik}(t)q_{1,j-k+1}.$$

记 $F_i(t,z)=\sum_{j=0}^{\infty}f_{ij}(t)z^j$，则 $i\geqslant 1$ 时

$$\begin{aligned}\frac{\partial F_i}{\partial t}&=\sum_{j=0}^{\infty}\Big(\sum_{k=1}^{j+1}kf_{ik}(t)q_{1,j-k+1}\Big)z^j\\&=\sum_{j=0}^{\infty}\Big(\sum_{k=0}^{j}(k+1)f_{i,k+1}(t)q_{1,j-k}\Big)z^j\\&=\Big(\sum_{j=0}^{\infty}(j+1)f_{i,j+1}(t)z^j\Big)\Big(\sum_{j=0}^{\infty}q_{1j}z^j\Big)=f(z)\frac{\partial F_i}{\partial z},\end{aligned}$$

因此 F_i 为下列方程的唯一解：

$$\frac{\partial F_i}{\partial t}=f(z)\frac{\partial F_i}{\partial z},\qquad F_i(0,z)=z^i. \tag{21}$$

另一方面，$(F_1)^i$ 也满足方程(21). 事实上，

$$\frac{\partial}{\partial t}(F_1)^i=i(F_1)^{i-1}\frac{\partial F_1}{\partial t}=i(F_1)^{i-1}f(z)\frac{\partial F_1}{\partial z}=f(z)\frac{\partial(F_1)^i}{\partial z}.$$

因此 $i\geqslant 1$ 时，$F_i(t,z)=[f_1(t,z)]^i$. $i=0$ 时显然有 $F_0(t,z)=p_{00}(t)=1=[F_1(t,z)]^0$. 利用科尔莫戈罗夫－查普曼方程也有 $F_1(t+s,z)=F_1(t,F_1(s,z))$，$t,s\geqslant 0$. 照搬前面的证明，我们

仍有 $\int_{z}^{F_1(t,z)} \frac{\mathrm{d}x}{f(x)} = t$. 若 $F_1(t,1)<1$,在这积分式中令 $z \uparrow 1$,就与条件(20)相矛盾了. 因此,$\sum_{j=0}^{\infty} f_{ij}(t) = F_1(t,1) = 1$,即最小解为转移概率函数,其密度矩阵是规则的,且是分支过程.

总之,保守马尔可夫链为分支过程的充要条件是它的密度矩阵为分支矩阵,且满足条件(20),这时链还是规则的. 利用(19)式可由密度矩阵决定分支过程的转移概率函数.

保守分支过程的跳跃阵为:

$$r_{00}=1, \quad r_{ij}=\begin{cases}0, & j<i-1, \\ h_{j-i+1}, & j \geqslant i-1,\end{cases} \quad i \geqslant 1,$$

其中

$$h_j=\begin{cases}0, & j=1, \\ \dfrac{q_{1j}}{q_1}, & j \neq 1.\end{cases}$$

由于 $r_{i,i-1}=h_0=q_{10}/q_1>0, i\geqslant 1$,对任意的 $i \geqslant 1, i \to 0$,但 0 是吸收状态. 我们同样要讨论吸收概率问题. 从状态 $i \geqslant 1$ 出发,被 0 吸收的概率为

$$f_{i0}=\lim_{t \to \infty} p_{i0}(t).$$

由(17)式,$p_{i0}(t)=[p_{10}(t)]^i$,因此 $f_{i0}=(f_{10})^i$. f_{10}也称为灭种概率. 下证 f_{10}即方程 $f(x)=0$ 的最小正根 σ. 我们已指出,或有 $\sigma=1$,或有 $0<\sigma<1$. 由于 $f(0)>0$,在$(0,\sigma)$上也有 $f(x)>0$. 此外必有

$$\int_0^{\sigma} \frac{\mathrm{d}x}{f(x)}=\infty, \tag{22}$$

在 $\sigma=1$ 时,这由条件(20)即得. 在 $\sigma<1$ 时,我们有

$$f(x)=f'(\sigma)(x-\sigma)+o(x-\sigma),$$

(22)也成立. 在(19)式中取 $z=0$,

$$\int_0^{G_t(0)} \frac{\mathrm{d}x}{f(x)}=t,$$

由(22)式可知,$t \to \infty$ 时 $G_t(0)$ 即 $p_{10}(t)$ 单调上升收敛于 σ. 所以,$f_{10}=\sigma$. 方程 $f(x)=0$ 即 $\sum_{j=0}^{\infty} h_j x^j = x$,由此可知,灭种概率 $f_{10}=1$ 的充要条件为

$$\sum_{j=1}^{\infty} j h_j \leqslant 1. \qquad \square$$

习　　题

3-1. 设在生灭过程中

$$\lambda_n=n^{\alpha}\lambda, n \geqslant 0; \quad \mu_n=n^{\alpha}\mu, n \geqslant 1, \quad \alpha,\lambda,\mu>0.$$

证明:过程为规则的充要条件为 $\alpha\leqslant 1$ 或 $\alpha>1,\lambda\leqslant\mu$.

3-2. 设在纯生过程中 $\lambda_n=n\lambda+\delta,n\geqslant 0,\lambda,\delta>0$,则

$$p_{ij}(t)=\begin{cases}0, & j<i,\\ e^{-(i\lambda+\delta)t}, & j=i,\\ \dfrac{1}{(j-i)!}\prod_{v=0}^{j-1}\left(i+\dfrac{\delta}{\lambda}+v\right)e^{-(i\lambda+\delta)t}(1-e^{-\lambda t})^{j-i}, & j>i.\end{cases}$$

3-3. 证明:纯灭过程的转移概率函数满足关系式:

$$p_{ij}(t)=0,\qquad j>i,$$
$$p_{ii}(t)=e^{-\mu_i t},$$
$$p_{ij}(t)=\mu_{j+1}e^{-\mu_j t}\int_0^t e^{\mu_j s}p_{i,j+1}(s)\mathrm{d}s,\qquad j<i.$$

特别,对线性纯灭过程:$\mu_n=n\mu,n\geqslant 1,\mu>0$,

$$p_{ij}(t)=C_i^j e^{-j\mu t}(1-e^{-\mu t})^{i-j},j\leqslant i.$$

3-4. 证明:在纯灭过程中,从状态 $i>0$ 出发到被状态 0 吸收(灭种)的平均时间为 $\dfrac{1}{\mu_1}+\dfrac{1}{\mu_2}+\cdots+\dfrac{1}{\mu_i}$.

3-5. 设 a 个人中有某一个人生某种传染病.在 Δt 时间内有两个人接触的概率为 $\alpha\Delta t$,多于一次接触的概率为 $o(\Delta t)$,任何两人接触的可能性完全相同.一个病人与一个非病人接触,则非病人被传染,且病人永远是病人.证明:全部 a 个人被传染的平均时间为 $\dfrac{a-1}{\alpha}\left(1+\dfrac{1}{2}+\cdots+\dfrac{1}{a-1}\right)$.

3-6. 证明:线性生灭过程的转移概率函数的母函数为

$$\sum_{n=0}^{\infty}p_{ij}(t)z^j=\begin{cases}\left\{\dfrac{\mu[1-e^{(\lambda-\mu)t}]-z[\lambda-\mu e^{(\lambda-\mu)t}]}{[\mu-\lambda e^{(\lambda-\mu)t}]-\lambda z[1-e^{(\lambda-\mu)t}]}\right\}^i, & \lambda\neq\mu,\\ \left[\dfrac{\lambda t+(1-\lambda t)z}{(1+\lambda t)-\lambda tz}\right]^i, & \lambda=\mu.\end{cases}$$

3-7. 设保守马尔可夫链的密度矩阵满足条件:$q_i\equiv\lambda>0$,则转移概率函数为

$$p_{ij}(t)=\sum_{n=0}^{\infty}r_{ij}^{(n)}\frac{(\lambda t)^n}{n!}e^{-\lambda t},$$

其中 $r_{ij}^{(n)}$ 为跳跃链的 n 阶转移概率.试给上述表示式一概率解释.

更一般地,若 $\sup_i q_i\leqslant\lambda<\infty$,则

$$p_{ij}(t)=\sum_{n=0}^{\infty}k_{ij}^{(n)}\frac{(\lambda t)^n}{n!}e^{-\lambda t},$$

其中$(k_{ij}^{(n)})=\boldsymbol{K}^n, \boldsymbol{K}=(k_{ij})$:

$$k_{ij}=\left(1-\frac{q_i}{\lambda}\right)\delta_{ij}+\frac{q_{ij}}{\lambda}(1-\delta_{ij}).$$

3-8. 证明:生灭过程为分支过程的充要条件是它为线性生灭过程:

$$l_n=n\lambda,\quad n\geqslant 0;\quad \mu_n=n\mu, n\geqslant 1, \lambda>0, \mu>0.$$

这时灭种概率为 $\min(1,\mu/\lambda)$.

3-9. 若分支矩阵满足条件:$\sum_{k=1}^{\infty} kq_{1k}<\infty$,则它是规则的.

3-10. 证明:连续时间分支过程的任一离散骨架是离散时间的分支过程,且每一离散骨架的灭种概率都与原来的连续时间分支过程的灭种概率相等.

§3.4 状态分类与平稳分布

与离散时间的马尔可夫链一样,我们对连续时间的马尔可夫链也要讨论状态的分类,转移概率函数的极限性质以及平稳分布.有些概念,例如两个状态的相通及在此基础上所作的状态分类等,可以直接从离散时间情形搬到连续时间情形;但也有些概念,例如常返性及正常返性等,却无法立即照搬过去.为了使我们的讨论只需用初等的方法,我们将通过对离散骨架的状态分类的讨论,建立起连续时间马尔可夫链的状态分类.这样做可以充分利用离散时间情形的已知结果.

定义 若存在 $t>0$,使 $p_{ij}(t)>0$,则称由状态 i 可达状态 j,记为 $i\to j$.若对一切 $t>0$,$p_{ij}(t)=0$,则称由状态 i 不可达状态 j,记为 $i\not\to j$.由定理 1.3,对一切 $t>0$,$p_{ii}(t)>0$,因此由状态 i 总可达状态 i,这点与离散时间情形有所不同.若 $i\to j$ 且 $j\to i$,则称状态 i 与 j 相通,记为 $i\leftrightarrow j$.完全类似地,可达与相通关系有传递性,即若 $i\to j, j\to k$,则 $i\to k$,或若 $i\leftrightarrow j, j\leftrightarrow k$,则 $i\leftrightarrow k$.从而可以按相通关系把状态分类,两两相通的状态组成一个状态类.状态 i 称为本质的,若 $i\to j$ 时,必有 $j\to i$.若存在 j 使得 $i\to j$,但 $j\not\to i$,则称状态 i 为非本质的.若整个状态空间是一个状态类,则称马尔可夫链是不可约的.

上面这些定义与离散时间情形是完全相同的.建立在可达与相通关系上的其它概念(如闭集等)和结果都可以搬过来使用,继续有效,我们不再重复叙述了.

定理 4.1 下列命题等价:

(1) 由状态 i 可达状态 j;

(2) 对任意的 h 骨架,由 i 可达 j;

(3) 对某一个 h 骨架,由 i 可达 j(记为 $i\xrightarrow{h} j$).

证 (2)⇒(3)⇒(1)是显然的,只要证(1)⇒(2).设 $t>0$ 使 $p_{ij}(t)>0$.取 n 充分大使 $nh>t$,则由定理 1.3 可知 $p_{ij}(nh)>0$,即 $i\xrightarrow{h} j$.这对一切 $h>0$ 成立. □

在定理 4.1 中,将可达改为相通显然也是成立的.定理 4.1 说明,从一个状态是否可达另一个状态,两个状态是否相通,对全部离散骨架和原来的连续时间马尔可夫脸都是一致的,从而它们的状态类也都是一致的,一个状态(或类)是本质或非本质也都是一致的.很自然地,我们希望一个状态是常返的或非常返的,正常返的或零常返的,也应该对全部离散骨架都是一致的.另一方面,离散时间情形定义状态为常返的或非常返的,正常返的或零常返的方法却不能直接搬到连

续时间情形,这也迫使我们试图从离散骨架的状态性质去寻找解决问题的途径.这些考虑导致下面的结果.

定理 4.2 对任意的 $i\in E$,若

$$\int_0^\infty p_{ii}(t)\mathrm{d}t=\infty, \tag{1}$$

则对一切 $h>0$,

$$\sum_n p_{ii}(nh)=\infty. \tag{2}$$

反之,若对某一个 $h>0$,(2)式成立,则(1)式成立.

证 $p_{ii}(h)$是$[0,\infty)$上严格正的连续函数,对任意的 $h>0$,

$$\begin{gathered}\min_{0\leqslant r\leqslant h}p_{ii}(r)=\delta(h)>0,\\ \min_{0\leqslant r\leqslant h}p_{ii}(t+r)\geqslant p_{ii}(t)\min_{0\leqslant r\leqslant h}p_{ii}(r)=p_{ii}(t)\delta(h),\\ m_n(h)=\min_{nh\leqslant t\leqslant(n+1)h}p_{ii}(t)\geqslant p_{ii}(nh)\delta(h).\end{gathered} \tag{3}$$

由(3)式,$0\leqslant r\leqslant h$ 时,

$$\begin{aligned}&p_{ii}(t)\geqslant p_{ii}(t-r)\delta(h),\\ &p_{ii}((n+1)h)\geqslant p_{ii}(nh+(h-r))\delta(h),\end{aligned}$$

从而有

$$\begin{aligned}p_{ii}((n+1)h)&\geqslant\max_{nh\leqslant t\leqslant(n+1)h}p_{ii}(t)\delta(h),\\ M_n(h)=\max_{nh\leqslant t\leqslant(n+1)h}p_{ii}(t)&\leqslant\frac{1}{\delta(h)}p_{ii}((n+1)h),\\ h\delta(h)\sum_{n=0}^{N-1}p_{ii}(nh)\leqslant h\sum_{n=0}^{N-1}m_n(h)&\leqslant\int_0^{Nh}p_{ii}(t)\mathrm{d}t\\ \leqslant h\sum_{n=0}^{N-1}M_n(h)&\leqslant\frac{h}{\delta(h)}\sum_{n=1}^{N}p_{ii}(nh),\end{aligned}$$

由此即得定理结论. □

由定理 4.2 可知,一个状态是常返的或非常返的,对全部的离散骨架来说是一致的.这就自然地引出下面的定义.

定义 对连续时间马尔可夫链,一个状态称为常返的或非常返的,若对全部离散骨架这个状态是常返的或非常返的.

定理 4.2 也提供了常返的判别法:状态 i 为常返的充要条件是

$$\int_0^\infty p_{ii}(t)\mathrm{d}t=\infty.$$

为了进一步讨论正常返与零常返,我们也要先讨论转移概率函数的极限性质.在离散时间情形,由于状态可能是周期的,因而使情况复杂化了.但在连续时间情形,由于对一切 $h>0$ 及正整数 n,对一切 $i\in E$,$p_{ii}(nh)>0$,对任一离散骨架,每个状态都是非周期的.所以,周期的概念对

连续时间的马尔可夫链就不需要了，由周期性造成的麻烦也就没有了.

定理 4.3 对一切 $i,j\in E$，存在极限

$$\lim_{t\to\infty} p_{ij}(t)=\pi_{ij}. \tag{4}$$

证 由离散时间情形的结果我们知道，对一切 $h>0$ 存在极限 $\lim_{n\to\infty} p_{ij}(nh)$. 我们要利用这个事实证明极限 $\lim_{t\to\infty} p_{ij}(t)$存在.

对任给的 $\varepsilon>0$，由于 $p_{ij}(t)$在$[0,\infty)$上一致连续，故有 $h>0$，使$|t-t'|<h$ 时

$$|p_{ij}(t)-p_{ij}(t')|<\frac{\varepsilon}{3}.$$

对这个 $h>0$，有正整数 N，使 $n,n'\geqslant N$ 时，

$$|p_{ij}(nh)-p_{ij}(n'h)|<\frac{\varepsilon}{3}.$$

这样，对任意的 $t,t'>Nh$，取正整数 n,n'，使得 $nh\leqslant t<(n+1)h$，$n'h\leqslant t'<(n'+1)h$，那么必定有 $n,n'\geqslant N$，$|t-nh|<h$，$|t'-n'h|<h$，因此

$$\begin{aligned}|p_{ij}(t)-p_{ij}(t')|&\leqslant|p_{ij}(t)-p_{ij}(nh)|+|p_{ij}(nh)-p_{ij}(n'h)|\\&\quad+|p_{ij}(n'h)-p_{ij}(t')|<\frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3}=\varepsilon,\end{aligned}$$

这即证明了极限 $\lim_{t\to\infty} p_{ij}(t)$存在. □

由定理 4.3，对一切 $h>0$ 极限

$$\lim_{n\to\infty} p_{ii}(nh)=\pi_{ii}=\pi_i,$$

与 h 无关，因此一个常返状态 i 为正常返的($\pi_i>0$)或零常返的($\pi_i=0$)对全部离散骨架来说也是一致的. 类似地给出下面的定义.

定义 对连续时间马尔可夫链，一个常返状态称为正常返的或零常返的，若对全部离散骨架这个状态是正常返的或零常返的.

状态 i 为正常返的充要条件是$\pi_i>0$.

对于有限状态的连续时间马尔可夫链，离散时间情形的结论显然继续成立：本质状态必为正常返状态；正常返状态必然存在；从非本质状态出发必定要进入正常返状态；有限不可约链的状态空间是一个正常返类.

现在可以将不变测度与平稳分布的概念及有关结果直接搬到连续时间情形了.

定义 非负数列$\{u_i,i\in E\}$称为连续时间马尔可夫链的不变测度，若对一切 $t>0$

$$u_i=\sum_k u_k p_{ki}(t),\qquad i\in E,$$

这也就是要求$\{u_i,i\in E\}$是全部离散骨架的不变测度. 若不变测度是概率分布，则称之为平稳分布.

由于 $\pi_i=\lim_{n\to\infty} p_{ii}(nh)$与 h 无关，因此对正常返不可约链，$\{\pi_i,i\in E\}$为平稳分布，进而离散时间情形平稳分布存在的条件及其一般形式对连续时间情形同样有效. 类似地，也有可配称

链与可逆链的概念.

定义 若存在不恒为零的非负数列 $\{u_i, i\in E\}$ 使得对一切 $t\geqslant 0, i, j\in E$,

$$u_i p_{ij}(t) = u_j p_{ji}(t),$$

则称马尔可夫链为可配称的.这时 $\{u_i, i\in E\}$ 也必为不变测度.若上述数列 $\{u_i, i\in E\}$ 又是概率分布,则称马尔可夫链为可逆的.这时 $\{u_i, i\in E\}$ 也必为平稳分布.

我们虽然已对连续时间马尔可夫链的状态分类作了合理的定义,但还没有给出有效的判别方法以及求平稳分布的方法,因为一般说来我们无法求得转移概率函数 $p_{ij}(t)$ 的明显表达式.我们早已指出,实际上能得到的是密度矩阵 (q_{ij}),然而只有在马尔可夫链是规则时,转移概率函数才由密度矩阵唯一决定.所以,下面我们只讨论规则链,找出根据密度矩阵判别状态性质及求平稳分布的方法.

定理 4.4 对规则链及 $i\neq j$, $i\to j$ 的充要条件是对其跳跃链,由 i 可达 j(记为 $i\xrightarrow{(r)}j$).

证 先证充分性.设 $i\xrightarrow{(r)}j$. i 不会是吸收状态(易见,对连续时间马尔可夫链及其跳跃链,吸收状态是一致的).若 $r_{ij}>0$,则 $q_{ij}>0$,然而 $q_{ij}=\lim_{t\to 0^+} p_{ij}(t)/t$, t 充分小时 $p_{ij}(t)>0$,从而 $i\to j$.一般地,存在互不相同的 $(i=)k_0, k_1, \cdots, k_n(=j)$,使得 $r_{k_l, k_{l+1}}>0, l=0,\cdots,n-1$,从而 $k_l\to k_{l+1}, l=0,\cdots,n-1$,即得 $i\to j$(值得注意的是,充分性的证明并没有用到规则性的假设).

再证必要性.设 $i\to j$,有 $t>0$,使 $p_{ij}(t)>0$.由规则性,并注意到 $i\neq j$, $p_{ij}(t)=f_{ij}(t)=\sum_{n=1}^{\infty} f_{ij}^{(n)}(t)$,因此必有 $n\geqslant 1$ 使 $f_{ij}^{(n)}(t)>0$.又由于

$$f_{ij}^{(n)}(t)=\sum_{k\neq i}\int_0^t \mathrm{e}^{-q_i(t-s)} q_{ik} f_{kj}^{(n-1)}(s)\mathrm{d}s>0,$$

必有 $k_1\neq i$ 使得

$$\int_0^t \mathrm{e}^{-q_i(t-s)} q_{ik} f_{k_1,j}^{(n-1)}(s)\mathrm{d}s>0,$$

从而 $q_{ik_1}>0$,并存在 $t_1\in(0,t)$ 使 $f_{k_1,j}^{(n-1)}(t_1)>0$.显然 $i\xrightarrow{(r)}k_1$,因此若 $k_1=j$,即有 $i\xrightarrow{(r)}j$.若 $k_1\neq j$,同理可知有 $k_2\neq k_1$,使 $q_{k_1k_2}>0$ $(k_1\xrightarrow{(r)}k_2)$ 及存在 $t_2\in(0,t_1)$ 使 $f_{k_2,j}^{(n-2)}(t_2)>0$.同样,若 $k_2=j$,即有 $i\xrightarrow{(r)}j$.若 $k_2\neq j$,则继续施以同样的手续,至多直到有 $k_{n-1}\neq k_{n-2}$,使 $q_{k_{n-2}k_{n-1}}>0$ $(k_{n-2}\xrightarrow{(r)}k_{n-1})$ 及存在 $t_{n-1}\in(0,t_{n-2})$ 使 $f_{k_{n-1},j}^{(1)}(t_{n-1})>0$,且 $k_{n-1}\neq j$.但这时

$$f_{k_{n-1},j}^{(1)}(t_{n-1})=\int_0^{t_{n-1}} \mathrm{e}^{-q_{k_{n-1}}(t-s)} q_{k_{n-1}j}\mathrm{e}^{-q_j s}\mathrm{d}s,$$

因此有 $q_{k_{n-1},j}>0$, $k_{n-1}\xrightarrow{(r)}j$,从而 $i\xrightarrow{(r)}k_1\xrightarrow{(r)}\cdots k_{n-1}\xrightarrow{(r)}j$.总之, $i\xrightarrow{(r)}j$ 成立. □

由定理 4.4 可知,规则链的状态类与它的跳跃链的状态类是一致的,一个类是本质的或非本质的对它们来说也是一致的.所以,只要判别各个本质类的状态性质.也不难看出,把一个规则链局限在一个本质类 C 上所得到的不可约链仍然是规则的.因为对新得到的不可约链,按定义写

出的最小解就是原来的最小解的一部分$\{f_{ij}(t), i,j\in C\}$,只要注意到 $i\in C, j\notin C$ 时 $q_{ij}=0$,从而对 $i\in C$, $\sum_{j\in C}f_{ij}(t)=1$.所以,我们只要判别规则不可约链的状态性质.

定理 4.5 规则不可约链为常返的充要条件是它的跳跃链是常返的.

证 显然,不可约链没有吸收状态,不然的话,状态空间只有一个状态,没有讨论的必要了.因此对一切 $i\in E, q_i>0$,我们要证明

$$\int_0^\infty p_{ij}(t)\mathrm{d}t=\frac{1}{q_j}\sum_{n=0}^\infty r_{ij}^{(n)}. \tag{5}$$

这里 $r_{ij}^{(n)}$是跳跃链的 n 步转移概率,由(5)及定理 4.2 即得结论.由于 $p_{ij}(t)=\sum_{n=0}^\infty f_{ij}^{(n)}(t)$,为了得到(5)式只需证明

$$\int_0^\infty f_{ij}^{(n)}(t)\mathrm{d}t=\frac{r_{ij}^{(n)}}{q_j}, \qquad n\geqslant 0. \tag{6}$$

现在用归纳法证(6)式. $n=0$ 时易算得:

$$\int_0^\infty f_{ij}^{(0)}(t)\mathrm{d}t=\delta_{ij}\int_0^\infty \mathrm{e}^{-q_it}\mathrm{d}t=\frac{\delta_{ij}}{q_i}=\frac{\delta_{ij}}{q_j}=\frac{r_{ij}^{(0)}}{q_j}.$$

设(6)式对 n 成立,则

$$\begin{aligned}\int_0^\infty f_{ij}^{(n+1)}(t)\mathrm{d}t &= \sum_{k\neq i}\int_0^\infty\left(\int_0^t \mathrm{e}^{-q_i(t-s)}q_{ik}f_{kj}^{(n)}(s)\mathrm{d}s\right)\mathrm{d}t\\ &=\sum_{k\neq i}q_{ik}\int_0^\infty f_{kj}^{(n)}(s)\mathrm{d}s\int_s^\infty \mathrm{e}^{-q_i(t-s)}\mathrm{d}t\\ &=\sum_{k\neq i}\frac{q_{ik}}{q_i}\int_0^\infty f_{kj}^{(n)}(s)\mathrm{d}s=\sum_k r_{ik}\frac{r_{kj}^{(n)}}{q_j}=\frac{r_{ij}^{(n+1)}}{q_j}.\end{aligned}$$ □

定理 4.6 规则不可约链为正常返的充要条件是方程组

$$\sum_k z_kq_{ki}=0, \quad i\in E \tag{7}$$

有非零非负收敛解.这时方程组(7)的非零非负收敛解与平稳分布只差一个常数.

证 先证必要性.这时$\{\pi_i, i\in E\}$为平稳分布:

$$\pi_i=\sum_k\pi_kp_{ki}(t), \quad t\geqslant 0, \quad i\in E.$$

取上式的拉普拉斯变换得

$$\frac{\pi_i}{\lambda}=\sum_k\pi_k\phi_{ki}(\lambda), \quad \lambda>0, \quad i\in E, \tag{8}$$

由向前方程我们有

$$(\lambda+q_i)\phi_{ij}(\lambda)=\delta_{ij}+\sum_{k\neq j}\phi_{ik}(\lambda)q_{kj},$$

将上式乘以 π_i,再对 i 相加得

$$(\lambda + q_j)\sum_i \pi_i \phi_{ij}(\lambda) = \pi_j + \sum_{k\neq j}\sum_i \pi_i\phi_{ik}(\lambda) q_{kj},$$

将(8)式代入上式

$$(\lambda + q_j)\frac{\pi_j}{\lambda} = \pi_j + \sum_{k\neq j}\frac{\pi_k}{\lambda} q_{kj},$$

整理一下即得

$$\sum_k \pi_k q_{kj} = 0, \qquad j \in E.$$

因此，$\{\pi_i, i\in E\}$是方程组(7)的非零非负收敛解.

再证充分性.设$\{u_i, i\in E\}$是方程组(7)的非零非负收敛解.先用归纳法证明对一切 $n\geqslant 0$,

$$\sum_k u_k \tilde{f}_{kj}^{(n)}(t) \leqslant u_j, \qquad j\in E, \quad t\geqslant 0, \tag{9}$$

这里 $\tilde{f}_{kj}^{(n)}(t) = \sum_{v=0}^{n} f_{kj}^{(v)}(t)$,

$$\tilde{f}_{kj}^{(0)}(t) = \delta_{kj}\mathrm{e}^{-q_j t},$$

$$\tilde{f}_{kj}^{(n+1)}(t) = \delta_{kj}\mathrm{e}^{-q_j t} + \sum_{l\neq j}\int_0^t \tilde{f}_{kl}^{(n)}(s) q_{lj}\mathrm{e}^{-q_j(t-s)}\mathrm{d}s, \quad n\geqslant 0.$$

易见 $n=0$ 时(9)成立：

$$\sum_k u_k \tilde{f}_{kj}^{(0)}(t) = u_j \mathrm{e}^{-q_j t} \leqslant u_j, \quad j\in E, \quad t\geqslant 0.$$

设(9)式对 n 成立，则

$$\begin{aligned}
\sum_k u_k \tilde{f}_{kj}^{(n+1)}(t) &= u_j\mathrm{e}^{-q_j t} + \sum_{l\neq j}\int_0^t \sum_k u_k \tilde{f}_{kl}^{(n)}(s) q_{lj}\mathrm{e}^{-q_j(t-s)}\mathrm{d}s \\
&\leqslant u_j\mathrm{e}^{-q_j t} + \int_0^t \sum_{l\neq j} u_l q_{lj}\mathrm{e}^{-q_j(t-s)}\mathrm{d}s \\
&\leqslant u_j\mathrm{e}^{-q_j t} + \int_0^t u_j q_j \mathrm{e}^{-q_j(t-s)}\mathrm{d}s = u_j.
\end{aligned}$$

在(9)式中取极限，由于 $\tilde{f}_{kj}^{(n)}(t)\uparrow p_{kj}(t)$，即得

$$\sum_k u_k p_{kj}(t) \leqslant u_j, \qquad j\in E, \quad t\geqslant 0.$$

再对 j 相加

$$\infty > \sum_k u_k \geqslant \sum_j\sum_k u_k p_{kj}(t) = \sum_k u_k \sum_j p_{kj}(t) = \sum_k u_k > 0,$$

因此必须成立等式

$$\sum_k u_k p_{kj}(t) = u_j, \qquad j\in E, \quad t\geqslant 0.$$

对离散骨架应用离散时间马尔可夫链的结果，即知链是正常返的，且 $u_i = (\sum_l u_j)\pi_i$，即$\{u_i, i\in E\}$与平稳分布只相差一个常数因子. □

定理 4.5 及定理 4.6 解决了判别规则马尔可夫链为常返、正常返及求平稳分布的方法.直观上从规则链的构造看,状态的转移状况与跳跃链的转移状况是完全一致的,因此可以预料常返性对两者是一致的.但正常返性涉及返回的时间长短,单由跳跃链不足以掌握原来的链在各个状态上停留的时间.很自然,一般来说正常返性对两者是不一致的.

例 4.1 规则不可约生灭过程 设生灭过程的密度矩阵为

$$Q=\begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

其中 $\lambda_i>0, i\geqslant 0, \mu_i>0, i\geqslant 1$.这时跳跃链是不可约的随机游动,其转移概率矩阵为:

$$\mathbf{R}=\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ \dfrac{\mu_1}{\lambda_1+\mu_1} & 0 & \dfrac{\lambda_1}{\lambda_1+\mu_1} & 0 & \cdots \\ 0 & \dfrac{\mu_2}{\lambda_2+\mu_2} & 0 & \dfrac{\lambda_2}{\lambda_2+\mu_2} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

因此,生灭过程也是不可约的.我们假设生灭过程是规则的,即满足条件

$$\frac{1}{\lambda_0}+\sum_{n=i}^{\infty}\left(\frac{1}{\lambda_n}+\frac{\mu_n}{\lambda_n\lambda_{n-1}}+\cdots+\frac{\mu_n\cdots\mu_1}{\lambda_n\cdots\lambda_0}\right)=\infty .$$

由定理 4.5 及第二章例 4.4,生灭过程为常返的充要条件是

$$\sum_{n=1}^{\infty}\frac{\mu_1\cdots\mu_n}{\lambda_1\cdots\lambda_n}=\infty .$$

为判别是否正常返,考察线性方程组(7):

$$\begin{cases} -\lambda_0 z_0+\mu_1 z_1=0, \\ \lambda_{n-1}z_{n-1}-(\lambda_n+\mu_n)z_n+\mu_{n+1}z_{n+1}=0, \qquad n\geqslant 1. \end{cases}$$

这方程组即为:

$$\mu_n z_n-\lambda_{n-1}z_{n-1}=0, \qquad n\geqslant 1,$$

$$z_n=\frac{\lambda_{n-1}}{\mu_n}z_{n-1}=\cdots=\frac{\lambda_{n-1}\cdots\lambda_0}{\mu_n\cdots\mu_1}z_0, \quad n\geqslant 1.$$

为使这方程组有非零非负收敛解,必须且只需

$$\mu_0=1+\sum_{n=1}^{\infty}\frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}<\infty ,$$

这也就是生灭过程为正常返的充要条件,且这时平稳分布为:

$$\pi_0=\frac{1}{\mu_0},\quad \pi_n=\frac{\lambda_0\cdots\lambda_{n-1}}{\mu_0\cdots\mu_n},\quad n\geqslant 1. \qquad \square$$

我们还可利用离散时间情形的相应结果,讨论常返不可约链的不变测度.

定理 4.7 常返不可约链有不变测度,且不计一个常数因子时不变测度是唯一的.

证 由于每个离散骨架都是常返不可约的,因此有非零非负数列$\{u_i,i\in E\}$(实际上每个$u_i>0$)使得

$$u_j=\sum_i u_ip_{ij}(1),\qquad j\in E.$$

我们要证明$\{u_i,i\in E\}$是不变测度,即对一切 $t>0$ 满足

$$u_j=\sum_i u_ip_{ij}(t),\qquad t\geqslant 0 \tag{10}$$

($t=0$ 时(10)式显然成立).先证 $t=n/m$ 为有理数时(10)式成立.

对$\frac{1}{m}$骨架,存在非零非负数列$\{\bar{u}_i,i\in E\}$使得对一切 $k\geqslant 1$

$$\bar{u}_j=\sum_i \bar{u}_ip_{ij}\left(\frac{k}{m}\right),\qquad j\in E. \tag{11}$$

取 $k=m$,则

$$\bar{u}_j=\sum_i \bar{u}_ip_{ij}(1),\qquad j\in E.$$

对1骨架应用离散时间情形的结果可知,$\{\bar{u}_i,i\in E\}$与$\{u_i,i\in E\}$只相差一个常数因子,因此(11)式中将$\{\bar{u}_i,i\in E\}$换成$\{u_i,i\in E\}$也成立.特别取 $k=n$,即知(10)式对 $t=n/m$ 成立.

对任意的 $t>0$,取一列有理数 $t_n\to t$,在

$$u_j=\sum_i u_ip_{ij}(t_n),\qquad j\in E$$

中令 $n\to\infty$ 即得

$$u_j\geqslant\sum_i u_ip_{ij}(t),\qquad j\in E.$$

若有一个 j 及 $t_0>0$ 使

$$u_j>\sum_i u_ip_{ij}(t_0),$$

取有理数 $r>t_0$,注意到 $p_{jj}(r-t_0)>0$,必须有

$$u_j=\sum_i u_ip_{ij}(r)=\sum_k\Big(\sum_i u_ip_{ik}(t_0)\Big)p_{kj}(r-t_0)<\sum_k u_kp_{kj}(r-t_0)\leqslant u_j,$$

得出矛盾,因此,对一切 $t>0$,(10)式成立. $\square$

定理 4.8 对规则常返不可约链,方程组

$$\sum_i z_iq_{ij}=0,\qquad j\in E \tag{12}$$

有非零非负解,且解即为不变测度.

证 由定理4.7,存在不变测度$\{u_i, i\in E\}$.照搬定理4.6的必要性的证明(那里只用到向前方程),可证$\{u_i, i\in E\}$为(12)的解.

现在设$\{u_i, i\in E\}$为(12)的非零非负解,我们只要证明$\{u_i, i\in E\}$是不变测度,即对一切$t>0$满足(10)式.仍照搬定理4.6中充分性的证明的前半部分,可得

$$\sum_i u_i p_{ij}(t)\leqslant u_j, \quad j\in E, \quad t>0,$$

我们要证明上式中全部成立等号.

首先指出,对一切$i\in E$, $u_i>0$.事实上,这时跳跃链是常返的不可约链,且由(12)式,

$$\sum_i (u_i q_i) r_{ij} = u_j q_j, \qquad j\in E$$

(我们不考虑状态空间仅由一个吸收状态组成的平凡情形),因此对一切i, $u_i q_i>0$, $u_i>0$.

其次,对任意的$t>0$,定义

$$\tilde{p}_{ij} = \frac{u_j}{u_i} p_{ji}(t), \qquad i,j\in E,$$

则$\tilde{p}_{ij}\geqslant 0$, $\sum_j \tilde{p}_{ij}\leqslant 1$.若有$i$使$\sum_j \tilde{p}_{ij}<1$,我们补充一个状态$\theta$,规定$\tilde{p}_{\theta\theta}=1$, $\tilde{p}_{i\theta}=1-\sum_j \tilde{p}_{ij}$, $i\in E$.这样,我们得到一个状态空间为$\tilde{E}=E\cup\{\theta\}$,以$(\tilde{p}_{ij})_{i,j\in\tilde{E}}$为转移概率矩阵的离散时间马尔可夫链.若状态$i\in E$使$\tilde{p}_{i\theta}>0$,则$i$在这个链中为非常返状态,$\sum_n \tilde{p}_{ii}^{(n)}<\infty$.但易用归纳法验证,$i,j\in E$时

$$\tilde{p}_{ij}^{(n)} = \frac{u_j}{u_i} p_{ji}(nt),$$

特别,$\tilde{p}_{ii}^{(n)} = p_{ii}(nt)$.因此,$\sum_n p_{ii}(nt)<\infty$.这与链为常返的相矛盾.所以只能对一切$i\in E$, $\sum_j \tilde{p}_{ij}=1$,即

$$\sum_j u_j p_{ji}(t) = u_i, \qquad i\in E. \qquad \square$$

定理4.9 规则马尔可夫链是可配称的充要条件为存在不全为零的非负数列$\{u_i, i\in E\}$,使得

$$u_i q_{ij} = u_j q_{ji}, \qquad i,j\in E. \tag{13}$$

证 必要性是容易的.因为这时存在不全为零的非负数列$\{u_i, i\in E\}$使得

$$u_i p_{ij}(t) = u_j p_{ji}(t), \qquad t\geqslant 0, \quad i,j\in E. \tag{14}$$

当$i\neq j$时,将(14)式在$t=0$点求导数即得(13).

充分性.现在存在不全为零的非负数列$\{u_i, i\in E\}$,使得(13)式成立.为证(14),只要证明$p_{ij}(t)$的拉普拉斯变换$\phi_{ij}(\lambda)$满足

$$u_i \phi_{ij}(\lambda) = u_i \phi_{ji}(\lambda), \qquad \lambda>0, \quad i,j\in E. \tag{15}$$

注意到,由定理 3.1 及 3.2,

$$\phi_{ij}(\lambda)=\sum_{n=0}^{\infty}\phi_{ij}^{(n)}(\lambda),$$

$$\phi_{ij}^{(0)}(\lambda)=\frac{\delta_{ij}}{\lambda+q_i}=\frac{\delta_{ij}}{\lambda+q_j},\tag{16}$$

$$\phi_{ij}^{(n+1)}(\lambda)=\sum_{k\neq i}\frac{q_{ik}}{\lambda+q_i}\phi_{kj}^{(n)}(\lambda)=\sum_{k\neq j}\phi_{ik}^{(n)}(\lambda)\frac{q_{kj}}{\lambda+q_j},\ n\geqslant 0.\tag{17}$$

因此,为证(15),只需证明,对一切 $n\geqslant 0$,

$$u_i\phi_{ij}^{(n)}(\lambda)=u_j\phi_{ji}^{(n)}(\lambda),\quad \lambda>0,\quad i,j\in E.\tag{18}$$

我们用归纳法证(18). $n=0$ 时,(18)由(16)即得:

$$u_i\phi_{ij}^{(0)}(\lambda)=\frac{u_i\delta_{ij}}{\lambda+q_i}=\frac{u_j\delta_{ij}}{\lambda+q_j}=u_j\phi_{ji}^{(0)}(\lambda).$$

设(18)式对 n 成立,则两次使用(17)式,得出

$$\begin{aligned}u_i\phi_{ij}^{(n+1)}(\lambda)&=\sum_{k\neq i}\frac{u_iq_{ik}\phi_{kj}^{(n)}(\lambda)}{\lambda+q_i}=\sum_{k\neq i}\frac{u_kq_{ki}\phi_{kj}^{(n)}(\lambda)}{\lambda+q_i}\\&=\sum_{k\neq i}\frac{q_{ki}u_j\phi_{jk}^{(n)}(\lambda)}{\lambda+q_i}=u_j\sum_{k\neq i}\frac{\phi_{jk}^{(n)}(\lambda)q_{ki}}{\lambda+q_i}=u_j\phi_{ji}^{(n+1)}(\lambda).\end{aligned}$$ □

例 4.2 规则生灭过程 在规则生灭过程中,设 $\mu_n>0,n\geqslant 1$,则过程是可配称的. 为此取

$$u_0=1,\qquad u_i=\frac{\lambda_{i-1}\cdots\lambda_0}{\mu_i\cdots\mu_1},\quad i\geqslant 1,$$

易见对一切 $i\geqslant 0$

$$u_iq_{i,i+1}=u_{i+1}q_{i+1,i}.$$

从而(13)式成立. 若又有 $\sum_i u_i<\infty$,则过程是可逆的. □

连续时间马尔可夫链在实际中有广泛的应用. 下面我们给出一些例子,着重表明连续时间马尔可夫链在运筹学(排队论、系统可靠性等)中的应用. 在这些例子中,连续时间马尔可夫链是适用的数学模型,对这些链的性质的讨论是必要的基础性工作. 在此基础上,就可按运筹的目标作出相应的决策. 这里主要讲述如何建立模型,并作基本的讨论. 关于如何决策的深入讨论请参阅运筹学的书籍.

例 4.3 排队系统 $M/M/S$ 在排队系统 $M/M/S$ 中,顾客来到的时间间隔服从指数分布,参数为 λ,服务时间也服从指数分布,参数为 μ,有 S 个服务员,按先来先服务的原则接待顾客.

我们以 X_t 表示时刻 t 系统中正在接受服务的及正在排队等候服务的顾客总数,即队伍长度. 在 $M/M/S$ 系统中,$\{X_t,t\geqslant 0\}$ 是一个连续时间的马尔可夫链. 事实上,如果已知现在系统中有 k 个顾客: $X_{t_0}=k$,则在将来的时刻 $t(>t_0)$ 的队伍长度 X_t,除了 k 之外,还取决于下列因素: (1) 现在正在接受服务的 $\min(k,S)$ 个顾客在 $(t_0,t]$ 内结束服务而离去的个数; (2) 在 $(t_0,t]$ 内新来到的顾客数; (3) 时刻 t_0 之后才开始接受服务的顾客在 $(t_0,t]$ 内结束服务而离去的个数. 由

于指数分布的无后效性,这三个因素都与时刻 t_0 之前的队伍长度无关.我们先计算密度矩阵.

若在时刻 t 有 i 个顾客在接受服务,则在 $(t,t+\Delta t]$ 内有一个顾客结束服务的概率为 $i\mu\Delta t + o(\Delta t)$,多于一个顾客服务完毕的概率为 $o(\Delta t)$.由于顾客的来到与正在接受服务的顾客结束服务的情况相互独立,因此在 $(t,t+\Delta t]$ 内既有顾客来到又有顾客结束服务而离去的概率为 $o(\Delta t)$.这样,密度矩阵 $\boldsymbol{Q}$ 为

$$\begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & \cdots & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & & \ddots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & S\mu & -(\lambda+S\mu) & \lambda & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & S\mu & -(\lambda+S\mu) & \lambda \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

从第 $(S+1)$ 行开始每一行都相同,因为至多只有 S 个顾客在接受服务.注意到对角线上的元素是有界的,这是一个规则不可约生灭过程.跳跃阵 $\boldsymbol{R}$ 为

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \frac{2\mu}{\lambda+2\mu} & 0 & \frac{\lambda}{\lambda+2\mu} & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & & \ddots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \frac{S\mu}{\lambda+S\mu} & 0 & \frac{\lambda}{\lambda+S\mu} & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & \frac{S\mu}{\lambda+S\mu} & 0 & \frac{\lambda}{\lambda+S\mu} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

记 $\rho=\lambda/S\mu$,应用例 4.1 的结果可得

(1) $\rho>1$ 时,链为非常返的;

(2) $\rho=1$ 时,链为零常返的;

(3) $\rho<1$ 时,链为正常返的,且这时平稳分布为

$$\pi_i=\begin{cases} \dfrac{1}{i!}\left(\dfrac{\lambda}{\mu}\right)^i\dfrac{1}{\mu_0}, & i=0,\cdots,S, \\ \dfrac{S^S}{S!}\left(\dfrac{\lambda}{S\mu}\right)^i\dfrac{1}{\mu_0}, & i=S+1,\cdots, \end{cases}$$

其中

$$\mu_0=\sum_{i=0}^{S}\frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i+\sum_{i=S+1}^{\infty}\frac{S^S}{S!}\left(\frac{\lambda}{S\mu}\right)^i.$$

$\rho\geqslant1$ 时,链为非正常返的,对一切 i,$\pi_i=0$.这表明队伍长度将无限增加,趋于无穷.$\rho<1$

时,时间充分长之后,队伍长度的分布趋于稳定:队长为 i 的概率为 π_i. $\rho=\lambda/S\mu$ 称为来往强度,它是反映系统稳定性的重要指标.从直观上看,λ 是单位时间内来到的顾客平均数,而 $S\mu$ 可看作(全部服务员都工作时)单位时间内能服务完毕的顾客平均数.因为 S 个相互独立的同指数分布(参数为 μ)的随机变量的最小值服从参数为 $S\mu$ 的指数分布,S 个服务员都服务时,服务完一个顾客的平均时间为 $1/S\mu$,它的倒数就是单位时间内服务完毕的顾客数.所以 $\rho>1$,即 $\lambda>S\mu$ 时,越来越多的顾客要排队等候,得不到服务,队伍长度要趋于无穷. $\rho<1$ 时,服务能力足以应付来到的顾客,系统是稳定的.只是 $\rho=1$ 的情形直观上较难判断.

在 $S=1$,即单个服务员的情况下,$\rho=\lambda/\mu<1$ 时平稳分布是几何分布:

$$\pi_i=(1-\rho)\rho^i, \qquad i\geqslant 0.$$

此时,一个顾客到达服务站,服务员空着立刻能得到服务的概率为 $\pi_0=1-\rho$,而遇到不少于 n 个顾客在服务站中的概率为 ρ^n.因此来往强度较小时系统遇到长的队伍的可能性是很小的. □

例 4.4 消失系统 如果在 $M/M/S$ 系统中没有地方给顾客排队等待,当一个顾客来到服务站,发现全部服务员都忙着服务,没有一个空着时,他就立刻离去,顾客"消失"了.这样的系统称为消失系统.在电话线路中就遇到这种情况.假设某电话总机总共有 S 条外线,来到的顾客就是分机要求接通外线的呼唤.如果这时有空着的外线,就可占用空着的线路通话,通话的时间就是服务时间.如果 S 条线路都已占满,那么要求通外线的用户只能挂断电话,呼唤即"消失"了.显然,在消失系统中没有在等待的顾客,这时队长过程 X_t 表示时刻 t 正在接受服务的顾客数(或正在服务的服务员个数).对情况作类似的分析表明 $\{X_t, t\geqslant 0\}$ 仍然是一个马尔可夫链,它有 $S+1$ 个状态:$\{0,1,\cdots,S\}$,密度矩阵 Q 为

$$\begin{pmatrix} \lambda & \lambda & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & & \ddots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots & (S-1)\mu & -[\lambda+(S-1)\mu] & \lambda \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & S\mu & -S\mu \end{pmatrix}.$$

这是一个有限不可约生灭过程,从而必定是正常返的,存在平稳分布.由例 4.1 的结果可知,平稳分布为

$$\pi_i=\frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i\Big/\sum_{k=0}^{S}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k, \qquad i=0,1,\cdots,S.$$

这就是电话工程中有名的爱尔朗(Erlang)公式.在系统运行充分长时间之后,可以认为在任一时刻,占用 i 条线路的概率为 π_i,全部线路被占用的概率为

$$\pi_S=\frac{1}{S!}\left(\frac{\lambda}{\mu}\right)^S\Big/\sum_{k=0}^{S}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k.$$

这也就是用户接不通外线,电话呼唤随即消失的概率,因此也称为消失概率.消失概率大,电话难打通,用户不满意,要求增加外线线路;增加了较多的线路,电话容易打通,但空闲的线路数也增

加了,使用效率低,经济上不合算.因此在设计线路时,要按照爱尔朗公式,综合考虑各种指标和费用,才能确定线路的条数.所以,爱尔朗公式在电话工程中起重要的作用. □

例 4.5　排队系统 $M/M/\infty$　在排队系统 $M/M/S$ 中,改变只有 S 个服务员的假定,假设有无穷多个服务员,因此顾客随到随服务,毋需等待,这样的系统记作 $M/M/\infty$.初看起来,有无穷多个服务员的假定似乎是不现实的,然而这个假定的实质是顾客随到随服务.假如那些由顾客自我服务的部门就是如此,从这个意义上说,有无穷多个服务员就是没有服务员.不仅在排队现象中可以遇到 $M/M/\infty$ 系统,在生物学家研究鸟类迁徙过程时也会碰到它.设某种鸟类飞入某地区的时间间隔服从指数分布.每只鸟在该地区停留的时间也服从指数分布.把停留时间看作服务时间,我们遇到的实际上就是 $M/M/\infty$ 系统,该地区中这种鸟的个数就是 $M/M/\infty$ 中的队伍长度.当然还可以有别的表面上很不同的现象归结为同一个系统.另一方面,当服务员个数 S 很大时,$M/M/\infty$ 也可作为 $M/M/S$ 的一种近似.大家知道,在数学上,讨论无限的情形往往比讨论有限的情形更方便,得到更简单的公式.

现在以 X_t 表示时刻 t 正在接受服务的顾客数,则 $\{X_t, t\geqslant 0\}$ 就是 $M/M/\infty$ 的队长过程,它同样也是一个马尔可夫链,密度矩阵 $\boldsymbol{Q}$ 为

$$\begin{pmatrix} \lambda & \lambda & 0 & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & 0 & \cdots \\ 0 & 0 & 3\mu & -(\lambda+3\mu) & \lambda & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

由定理 3.5 易见,这是一个规则密度矩阵.因此,$\{X_t, t\geqslant 0\}$ 仍是一个规则不可约生灭过程.由例 4.1 的结果,因 $\sum_n \frac{\lambda^n}{n!\ \mu^n}<\infty$,过程是正常返的,且平稳分布为

$$\pi_i=\frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i \mathrm{e}^{-\lambda/\mu},\qquad i\geqslant 0,$$

即参数为 λ/μ 的泊松分布.

对这个过程还可解向前方程,求出转移概率函数:

$$\begin{cases} p_{i0}'(t)=\lambda p_{i0}(t)+\mu p_{i1}(t), \\ p_{ij}'(t)=\lambda p_{i,j-1}(t)-(\lambda+j\mu)p_{ij}(t)+(j+1)\mu p_{ij}(t), \quad j\geqslant 1. \end{cases}$$

令 $\Phi_i(t,z)=\sum_{j=0}^{\infty} p_{ij}(t)z^j$,我们有

$$\frac{\partial\Phi_i}{\partial t}=\lambda(z-1)\Phi_i-\mu(z-1)\frac{\partial\Phi_i}{\partial z}.$$

作代换

$$\begin{cases} s=(z-1)\mathrm{e}^{-\mu t}, \\ u=\dfrac{\lambda}{\mu}z, \end{cases}$$

$$\frac{\partial\Phi_i}{\partial t}=\frac{\partial\Phi_i}{\partial s}\frac{\partial s}{\partial t}+\frac{\partial\Phi_i}{\partial u}\frac{\partial u}{\partial t}=-\mu(z-1)\mathrm{e}^{-\mu t}\frac{\partial\Phi_i}{\partial s}$$

$$\frac{\partial \Phi_i}{\partial z}=\frac{\partial \Phi_i}{\partial s}\frac{\partial s}{\partial z}+\frac{\partial \Phi_i}{\partial u}\frac{\partial u}{\partial z}=e^{-\mu t}\frac{\partial \Phi_i}{\partial s}+\frac{\lambda}{\mu}\frac{\partial \Phi_i}{\partial u},$$

偏微分方程化为最简单的形式：

$$\frac{\partial \Phi_i}{\partial u}=\Phi_C,$$

所以

$$\Phi_i(s,u)=e^u\Phi(s),$$

其中 Φ 为某待定的函数.回到变量 t 及 z：

$$\Phi(t,z)=e^{\lambda/\mu}\Phi[(z-1)e^{-\mu t}],$$

由初始条件 $\Phi_i(0,z)=z^i$，$z^i=e^{\frac{\lambda}{\mu}z}\Phi(z-1)$，可解得

$$\Phi(z)=(1+z)^i\exp\{-\frac{\lambda}{\mu}(1+z)\},$$

$$\Phi_i(t,z)=[1+(z-1)e^{-\mu t}]^i\exp\left\{\frac{\lambda}{\mu}(z-1)(1-e^{-\mu t})\right\}.$$

展开成幂级数，比较系数可得

$$p_{ij}(t)=\sum_{k=0}^{\min(i,j)}C_i^k\left(\frac{\lambda}{\mu}\right)^{j-k}\frac{e^{-k\mu t}(1-e^{-\mu t})^{i+j-2k}}{(j-k)!}\exp\left\{-\frac{\lambda}{\mu}(1-e^{-\mu t})\right\}.$$

特别，若 $X_0=0$，则

$$\mathrm{P}(X_i=j)=\frac{1}{j!}\left[\frac{\lambda}{\mu}(1-e^{-\mu t})\right]^j\exp\left\{-\frac{\lambda}{\mu}(1-e^{-\mu t})\right\},$$

也是泊松分布.在此式中令 $t\to\infty$ 即得平稳分布. □

例 4.6 机器修理问题 设有 M 台机床，S 个修理工，$S\leqslant M$.机床或者在工作，或者已损坏而等待修理.机床损坏后，如有修理工空着，那么立即对它修理，否则要等到有一个修理工有空之后再来修理.损坏的机床修理好即重新开始工作.假定时刻 t 正在工作的每台机床在 $(t,t+\Delta t)$ 中损坏的概率为 $\lambda\Delta t+o(\Delta t)$，即机床正常工作的寿命服从参数为 λ 的指数分布；时刻 t 正在修理的每台机床在 $(t,t+\Delta t)$ 中被修好的概率为 $\mu\Delta t+o(\Delta t)$，即修理时间服从参数为 μ 的指数分布；各机床的状况相互独立.

以 X_t 表示时刻 t 不在工作的机床数，根据上面的假定，$\{X_t,t\geqslant 0\}$ 是一个马尔可夫链，有 $M+1$ 状态：$\{0,1,\cdots,M\}$，密度矩阵 Q 为

$$\begin{pmatrix}
-M\lambda & M\lambda & 0 & 0 & \cdots & \cdots & 0\\
\mu & -[(M-1)\lambda+\mu] & (M-1)\lambda & 0 & \cdots & \cdots & 0\\
 & & \ddots & & & & \\
0 & \cdots & S\mu & -[(M-S)\lambda+S\mu] & (M-S)\lambda & \cdots & 0\\
\vdots & & & & \ddots & & \\
0 & \cdots & \cdots & S\mu & -(\lambda+S\mu) & \lambda & 0\\
0 & \cdots & \cdots & \cdots & 0 & S\mu & -S\mu
\end{pmatrix}$$

这是正常返的有限不可约生灭过程.平稳分布满足下列方程组:

$$\begin{cases}-M\lambda z_0+\mu z_1=0,\\(M-i+1)\lambda z_{i-1}-[(M-i)\lambda+i\mu]z_i+(i+1)\mu z_{i+1}=0, & i=1,\cdots,S-1,\\(M-i+1)\lambda z_{i-1}-[(M-i)\lambda+S\mu]z_i+S\mu z_{i+1}=0, & i=S,\cdots,M-1,\\\lambda z_{M-1}-S\mu z_M=0,\end{cases}$$

解得

$$\pi_i=\begin{cases}C_M^i\left(\dfrac{\lambda}{\mu}\right)^i\pi_0, & i=1,\cdots,S,\\C_M^i\dfrac{(S+1)\cdots i}{S^{i-S}}\left(\dfrac{\lambda}{\mu}\right)^i\pi_0, & i=S+1,\cdots,M,\end{cases}$$

$$\pi_0=\left\{\sum_{k=0}^{S}C_M^k\left(\frac{\lambda}{\mu}\right)^k+\sum_{k=S+1}^{M}C_M^k\frac{(S+1)\cdots k}{S^{k-S}}\left(\frac{\lambda}{\mu}\right)^k\right\}^{-1}.$$

在 $S=M$ 时,注意到 $\sum_{k=0}^{M}C_M^k\left(\frac{\lambda}{\mu}\right)^k=\left(1+\frac{\lambda}{\mu}\right)^M$,可得

$$\pi_i=C_M^i\left(\frac{\lambda}{\lambda+\mu}\right)^i\left(\frac{\mu}{\lambda+\mu}\right)^{M-i},\qquad i=0,\cdots,M,$$

即为二项分布.当然,M 个修理工应付 M 台机床是不经济的.怎样合理地配备修理工人要作仔细的分析.

$\sum_{k=1}^{M}k\pi_k$ 是不在工作的机床的平均台数,因此定义

$$\text{机床平均利用率}=1-\frac{1}{M}\sum_{k=1}^{M}k\pi_k.$$

当不工作的机床数 $k\leqslant S$ 时,修理工的利用率为 k/S;不工作的机床数超过 S 时,全部修理工都在修机床,利用率为 1,因此定义

$$\text{修理工平均利用率}=\sum_{k=1}^{S}\frac{k}{S}\pi_k+\sum_{k=S+1}^{M}\pi_k.$$

下表提供的数字有助于我们作进一步的分析.

M \ ρ	0.1	0.5	1
5	0.872 1/0.715 1	0.385 3/0.963 3	0.199 4/0.996 9
10	0.785 4/0.981 6	0.2/1	0.1/1
15	0.642 3/0.999 8	0.133 3/1	0.066 7/1

$S=1,\rho=\lambda/\mu$　机床平均利用率/修理工平均利用率

从上表中可看出,机床平均利用率与修理工平均利用率是相互矛盾的.通常在 ρ 确定之后(ρ 取决于机床及修理工的水平),应根据各项指标与费用适当地决定配备修理工的数量.

M	S	修理工平均利用率
4	1	0.881
8	2	0.934
16	4	0.994

$\rho=0.45$

这个表提供的数据是很有启发性的.尽管修理工与机床数的比例保持不变,但随着机床数的增加,修理工平均利用率有明显的提高.所以修理工的配备方式也是重要的.分组配备不如统一使用效率高.这个道理直观上也是容易想到的,但我们提供了具体的计算方法,从而有确切的数据作依据.在修理工平均利用率接近于1时,再过分地集中使用修理工意义就不大了.我们的数据也为如何分组配备修理工提供了依据. □

例 4.7 系统可靠性问题 设一个系统仅由一台机器组成,它正常工作的时间服从参数为 λ 的指数分布的随机变量.机器损坏后即进行修理,修复的时间是服从参数为 μ 的指数分布的随机变量,修复后即正常工作.这实际上是机器修理问题的最简单的情形:一部机器及一个修理工.系统正常工作的概率为

$$\pi_0=\frac{\mu}{\lambda+\mu}.$$

称为系统的有效度.有效度有下述有趣的直观解释.机器平均正常工作的时间为 $1/\lambda$,平均修理时间为 $1/\mu$.因此,一个工作—修理的周期的平均长度为 $1/\lambda+1/\mu$,其中工作时间所占比例恰为 $(1/\lambda)/(1/\lambda+1/\mu)=\mu/(\lambda+\mu)$,即有效度.如果一个系统由 n 台机器组成,每台机器的运转状况同上,且各台机器的状况相互独立,那么这就是 n 台机器及 n 个修理工的机器修理问题.这时系统的有效度是 n 台机器全在工作的概率

$$\pi_0=\left(\frac{\mu}{\lambda+\mu}\right)^n.$$

它等于各台机器的有效度的乘积,显然这是独立性假设的结果.

现在设一个系统由 n 台不同性能的机器组成,第 i 台机器的工作时间服从参数为 λ_i 的指数分布,修理时间服从参数为 μ_i 的指数分布,各台机器的状况仍相互独立.我们仍要求系统的有效度,即全部机器都在工作的概率.为简单起见,我们只对 $n=2$ 作计算.

规定下列四种状态:“0”表示两台机器都在工作;“1”表示第一台机器损坏在修理,第二台在工作;“2”表示第一台在工作,第二台损坏在修理;“3”表示两台机器均损坏在修理.以 X_t 表示时刻 t 系统的状态,则 $\{X_t,t\geqslant 0\}$ 为具有上述四个状态的马尔可夫链,密度矩阵 Q 为

$$\begin{pmatrix} -(\lambda_1+\lambda_2) & \lambda_1 & \lambda_2 & 0 \\ \mu_1 & -(\mu_1+\lambda_2) & 0 & \lambda_2 \\ \mu_2 & 0 & -(\mu_2+\lambda_1) & \lambda_1 \\ 0 & \mu_2 & \mu_1 & -(\mu_1+\mu_2) \end{pmatrix},$$

平稳分布满足方程组：

$$\begin{cases}(\lambda_1+\lambda_2)z_0=\mu_1 z_1+\mu_2 z_2,\\(\mu_1+\lambda_2)z_1=\lambda_1 z_0+\mu_2 z_3,\\(\mu_2+\lambda_1)z_2=\lambda_2 z_0+\mu_1 z_3,\\(\mu_1+\mu_2)z_3=\lambda_2 z_1+\lambda_1 z_2.\end{cases}$$

易见方程组有解：

$$z_0=\mu_1\mu_2,\quad z_1=\lambda_1\mu_2,\quad z_2=\lambda_2\mu_1,\quad z_3=\lambda_1\lambda_2,$$

我们要求的有效度为

$$\pi_0=\frac{\mu_1}{\lambda_1+\mu_1}\frac{\mu_2}{\lambda_2+\mu_2}.$$

一般地，系统由 n 台机器组成时，有效度为

$$\pi_0=\prod_{i=1}^{n}\frac{\mu_i}{\lambda_i+\mu_i}.$$

它同样是各台机器的有效度的乘积，其中的道理是一样的. □

例 4.8　停工检修问题　设一个车间里有 n 部机器，每部机器正常工作的时间都服从参数为 λ 的指数分布，且各部机器的工作状况相互独立. 如果机器损坏了 k 部 $(1\leqslant k\leqslant m)$，为了进行修理，必须全车间停工，而且修理的时间服从参数为 μ 的指数分布，与 k 无关. 现在的问题是如何决定这个 k，当损坏了 k 部机器时就停工检修，使得全车间的生产率最高，即在工作的机器的平均台数最多.

令 X_t 为时刻 t 损坏的机器数. 由假设，$\{X_t,t\geqslant 0\}$ 为马尔可夫链，有 $k+1$ 个状态：$\{0,1,\cdots,k\}$，密度矩阵 Q 为

$$\begin{pmatrix}-m\lambda & m\lambda & 0 & \cdots & \cdots & 0\\ 0 & -(m-1)\lambda & (m-1)\lambda & \cdots & \cdots & 0\\ \cdots & \cdots & \ddots & & \cdots & \cdots\\ 0 & \cdots & \cdots & 0 & -(m-k+1)\lambda & (m-k+1)\lambda\\ \mu & 0 & 0 & \cdots & \cdots & -\mu\end{pmatrix}.$$

这里只需注意，在损坏了 k 部机器之后，经过修理即由状态 k 进入状态 0. 平稳分布满足方程组：

$$\begin{cases}-m\lambda z_0+\mu z_k=0,\\(m-i+1)\lambda z_{i-1}-(m-i)\lambda z_i=0,\quad i=1,\cdots,k-1,\\(m-k+1)\lambda z_{k-1}-\mu z_k=0,\end{cases}$$

由此解得：

$$\begin{cases}z_k=m\dfrac{\lambda}{\mu}z_0,\\z_i=\dfrac{m-i+1}{m-i}z_{i-1}=\cdots=\dfrac{m}{m-i}z_0,\quad i=1,\cdots,k-1,\end{cases}$$

$$\begin{cases}\pi_i = \dfrac{m}{m-i}\pi_0, \quad i=1,\cdots,k-1, \qquad \pi_k = \dfrac{m\lambda}{\mu}\pi_0, \\ \pi_0 = \left\{ m\left(\sum_{i=0}^{k-1}\dfrac{1}{m-i} + \dfrac{\lambda}{\mu}\right)\right\}^{-1}.\end{cases}$$

工作着的机器的平均台数为

$$\begin{aligned}A_k &= m\pi_0 + (m-1)\pi_1 + \cdots + (m-k+1)\pi_{k-1} \\ &= k\left\{\sum_{i=0}^{k-1}\frac{1}{m-i} + \frac{\lambda}{\mu}\right\}^{-1}.\end{aligned}$$

我们要选取 k,使 A_k 达到最大.

$\{A_k, k\geqslant 1\}$具有下列性质:若 $A_k\geqslant A_{k+1}$,则 $A_k > A_{k+\upsilon}$,$\upsilon\geqslant 2$.事实上,$A_k\geqslant A_{k+1}$即

$$\frac{\lambda}{\mu}\leqslant\frac{k}{m-k} - \sum_{i=0}^{k-1}\frac{1}{m-i}.$$

$\upsilon\geqslant 2$ 时

$$\begin{aligned}(k+\upsilon)\frac{\lambda}{\mu} + (k+\upsilon)\sum_{i=0}^{k-1}\frac{1}{m-i} &\leqslant k\frac{\lambda}{\mu} + k\left[\frac{\upsilon}{m-k} + \sum_{i=0}^{k-1}\frac{1}{m-i}\right] \\ &< k\sum_{i=0}^{k+\upsilon-1}\frac{1}{m-i} + k\frac{\lambda}{\mu},\end{aligned}$$

这即 $A_k > A_{k+\upsilon}$.由此可得下列结论:

若

$$\frac{\lambda}{\mu}\leqslant\frac{1}{m(m-1)},$$

即 $A_1\geqslant A_2$,则 $A_1 = \max_k A_k$,应取 $k=1$;

若

$$\frac{1}{m(m-1)} < \frac{\lambda}{\mu}\leqslant\frac{3m-2}{m(m-1)(m-2)},$$

即 $A_1 < A_2, A_2\geqslant A_3$,则 $A_2 = \max_k A_k$,应取 $k=2$;

一般地,若

$$\frac{j}{m-j} - \sum_{i=0}^{j-1}\frac{1}{m-i} < \frac{\lambda}{\mu}\leqslant\frac{j+1}{m-j-1} - \sum_{i=0}^{j}\frac{1}{m-i}, 1\leqslant j\leqslant m-2,$$

即 $A_1 < A_2 < \cdots < A_{j+1}\geqslant A_{j+2}$,则 $A_{j+1} = \max_k A_k$,应取 $k=j+1$;

最后,若

$$\frac{\lambda}{\mu} > (m-1) - \sum_{i=0}^{m-2}\frac{1}{m-i},$$

即 $A_1 < A_2 < \cdots < A_m$,则 $A_m = \max_k A_k$,应取 $k=m$. □

习　题

4－1. 证明：有 a 个状态的不可约链的密度矩阵的秩为 $a-1$.

4－2. 对规则不可约链，证明：

(1) 若 $\sup_i q_i \leqslant \infty$，且链为正常返的，则跳跃链也为正常返的；

(2) 若 $\inf_i q_i > 0$，且跳跃链为正常返的，则链为正常返的.

4－3. 设在生灭过程中 $\mu_n = n, n \geqslant 1$.

(1) 若 $\lambda_n = n+2, n \geqslant 0$，则链为非常返不可约的；

(2) 若 $\lambda_n = n+1, n \geqslant 0$，则链为零常返不可约的；

(3) 若 $\lambda_n = \max(n-1,1), n \geqslant 0$，则链为正常返不可约的，求其平稳分布.

4－4. 设在生灭过程中，$\lambda_n = \lambda > 0, n \geqslant 0, \mu_n = n, n \geqslant 1$，则链为正常返不可约的，且平稳分布是参数为 λ 的泊松分布. 若以平稳分布为初始分布，则 $(0,t]$ 内的死亡次数服从参数为 λt 的泊松分布.

4－5. 设在生灭过程中，$\lambda_n = n\lambda + \delta, n \geqslant 0, \lambda, \delta > 0, \mu_n = n\mu, n \geqslant 1, \mu > 0$，则

(1) $\lambda > \mu$ 或 $\lambda = \mu < \delta$ 时，链为非常返的；

(2) $\lambda = \mu \geqslant \delta$ 时，链为零常返的；

(3) $\lambda < \mu$ 时，链为正常返的，且这时平稳分布为

$$\begin{cases} \pi_0 = \left(1 - \dfrac{\lambda}{\mu}\right)^{\delta/\lambda}, \\ \pi_n = \dfrac{1}{n!}\dfrac{\delta}{\lambda}\left(\dfrac{\delta}{\lambda}+1\right)\cdots\left(\dfrac{\delta}{\lambda}+n-1\right)\left(\dfrac{\lambda}{\mu}\right)\left(1-\dfrac{\lambda}{\mu}\right)^{\delta/\lambda}, \quad n \geqslant 1. \end{cases}$$

4－6. 在排队系统 $M/M/1$ 中，当顾客来到时看见队长超过 $a(>1)$ 时，以概率 p 留下来参加排队，以概率 $1-p$ 离去$(0<p<1)$，则队长过程 $\{X_t, t \geqslant 0\}$ 为不可约马尔可夫链，试讨论它的性状.

4－7. 一个汽车加油站每次只能给一辆汽车加油，加油时间服从参数为 μ 的指数分布，各辆汽车的加油时间相互独立. 加油的汽车按参数为 λ 的泊松过程来到，但当一辆汽车来到加油站发现站中已有 n 辆汽车时，以概率 $\dfrac{n}{n+1}$ 立即离去，以概率 $\dfrac{1}{n+1}$ 留下来排队. 记 X_t 为时刻 t 加油站中汽车的数目，则 $\{X_t, t \geqslant 0\}$ 为正常返不可约马尔可夫链，求其平稳分布.

4－8. 设在排队系统 $M/M/1$ 中再添加一个服务员，但仅在队长超过 a 时，才让两个服务员参加服务，当队长降到 a 时就撤掉一个服务员. 记 X_t 为时刻 t 的队伍长度，则 $\{X_t, t \geqslant 0\}$ 为不可约马尔可夫链，试讨论它的性状.

4－9. 设在排队系统 $M/M/2$ 中，来往强度 $\rho = \lambda/2\mu < 1$. 现在用一个服务员代替原来的两个服务员，但新的服务员的服务速度比老服务员快一倍，即在新的 $M/M/1$ 系统中服务时间服从参数为 2μ 的指数分布. 试比较这两个系统的平均队长，以说明这个措施的效果如何.

4－10. 一个出租汽车服务站上，出租汽车按参数为 λ 的泊松过程到站候客，顾客按参数为 μ 的泊松过程到站候车，这两个过程相互独立. 当一个顾客到站时，看到已有两个顾客在候车时，就立即离去不参加排队(一辆车载一位顾客). 设 $\rho = \lambda/\mu < 1$，求候车的顾客平均数及候客的出

租汽车平均数.

4－11. 设顾客按参数为 λ 的泊松过程来到某服务部门. 每个顾客必须相继受两次服务，两次服务时间相互独立，且分别服从参数为 μ_1 及 μ_2 的指数分布. 只有在两个服务台均空闲时，才接纳顾客. 只要有一个服务台在工作，来到的顾客就立即离去. 求顾客来到时能受到服务的概率.

4－12. 某车间有 n 部车床，每部车床的电动机的功率为 a 千瓦. 一部车床有时开动着在车削，有时关闭电动机，停下来调换工件或进行测量. 假设在时刻 t 车床在车削，那么在 $(t, t+\Delta t)$ 内车床停下来的概率为 $\lambda\Delta t + o(\Delta t)$. 若在时刻 t 车床停着，那么在 $(t, t+\Delta t)$ 内车床又开动起来的概率为 $\mu\Delta t + o(\Delta t)$. 假设各部车床的工作状况相互独立，全车间平均消耗电力多少千瓦?

4－13. 在 M 部机器与 M 个修理工的机器修理问题中，以 X_t 记时刻 t 不在工作的机器数，则 $\{X_t, t \geqslant 0\}$ 为埃伦弗斯特扩散模型(参见习题 2－6.).

4－14. 一个系统由 n 个不同的部件串联而成. n 个部件的寿命分别服从参数为 λ_i 的指数分布，失效后的修理时间分别服从参数为 μ_i 的指数分布. 若 n 个部件都正常工作，则系统处于工作状态. 若有某个部件失效，则系统失效. 这时修理工立即对失效部件进行修理，其余部件停止工作. 若失效部件修复，所有部件立即进入工作状态，从而系统重新处于工作状态. 假设各部件的状况相互独立，求系统的有效度(系统处于工作状态的概率).

4－15. 一个系统由 n 个相同部件并联而成. n 个部件的寿命都服从参数为 λ 的指数分布，失效后的修理时间服从参数为 μ 的指数分布. 只要有一个部件在工作，系统就处于工作状态；全部部件失效时，系统才失效. 假设只有一个修理工，当修理工在修理一个失效部件时，其它失效部件必须等待被修理. 假设各部件的状况相互独立，求系统失效的概率.

4－16. 一个冷贮备系统由 n 个相同部件及一个修理工组成. 开始时一个部件工作，其余部件不工作(冷贮备). 若工作部件失效，贮备部件之一立即去顶替，转为工作状态，同时修理工立即对失效部件进行修理. 当修理工在修理某一失效部件时，其它的失效部件必须等待被修理. 修好的部件或立即投入工作(如其它部件均已失效)，或进入冷贮备状态(如已有部件在工作). 系统在全部部件失效时失效. 设 n 个部件的工作寿命都服从参数为 λ 的指数分布，失效后的修理时间服从参数为 μ 的指数分布，且各个部件的状况相互独立，求系统失效的概率.

4－17. 一个冷贮备系统由两个部件及一个修理工组成，其中部件甲比部件乙有优先权. 优先权的含义是：(1) 当失效的部件甲被修好时，若部件乙正在工作，则部件甲立即转为工作状态，部件乙停止工作，转入贮备；(2) 当部件甲失效时，若部件乙正在被修理，则停止修理部件乙，修理工转而先修理甲. 部件甲与乙的工作寿命分别服从参数为 λ_1 与 λ_2 的指数分布，修理时间分别服从参数为 μ_1 与 μ_2 的指数分布. 假设两个部件的状况相互独立，求系统失效的概率.

4－18. 一个热贮备系统由两个相同部件及一修理工组成. 开始时两个部件都正常，一个处于工作状态，一个处于热贮备状态. 处于工作状态的部件的寿命服从参数为 λ 的指数分布；处于热贮备状态的部件的寿命服从参数为 υ 的指数分布. 若贮备部件先失效，立即修理失效的部件，工作部件继续工作. 若工作部件先失效，失效部件也立即得到修理，同时贮备部件转为工作部件，其工作寿命也转为服从参数为 λ 的指数分布，与它已贮备的时间无关. 两个部件都失效(一个在修理，另一个等待被修理)，则系统失效. 假定两个部件的修理时间都服从参数为 μ 的指数分布，且两个部件的状况相互独立，求系统失效的概率.

第四章　更新过程

§4.1　更新过程与更新方程

定义　设$\{W_n, n\geqslant 1\}$为i.i.d.随机变量序列，$W_n\geqslant 0$，其共同分布

$$F(t)=\mathsf{P}(W_1\leqslant t)$$

(注意，我们使用右连续的分布函数)满足条件

$$F(0)=\mathsf{P}(W_1=0)<1. \tag{1}$$

令

$$S_0=0,\quad S_n=\sum_{j=1}^{n} W_j,\quad n\geqslant 1,$$

$$N(t)=\sup\{n: S_n\leqslant t\}=\sum_{n=1}^{\infty} 1_{\{S_n\leqslant t\}},$$

则计数过程$N=\{N(t), t\geqslant 0\}$称为更新过程. W_n称为N的第n个更新间隔(跳跃间隔)或第n次更新的等待时间，S_n称为第n次更新时刻，W_n的分布F称为更新过程的等待时间分布.

由定义式可知，对一切$n\geqslant 1$及$t\geqslant 0$

$$\{N(t)\geqslant n\}=\{S_n\leqslant t\}, \tag{2}$$

$$\{N(t)=n\}=\{S_n\leqslant t<S_{n+1}\}, \tag{3}$$

$$N(t)+1=\inf\{n\geqslant 1: S_n>t\}. \tag{4}$$

因此，$N(t)+1$是关于$\{W_n, n\geqslant 1\}$或$\{S_n, n\geqslant 0\}$的停时.

由强大数律，以概率1有

$$\frac{1}{n}S_n=\frac{1}{n}\sum_{j=1}^{n} W_j\to\mu=\mathsf{E}\,W_1.$$

条件(1)保证$\mu>0$(但可能$\mu=\infty$)，因而以概率1有$S_n\to\infty$，对一切$t\geqslant 0$

$$\mathsf{P}(N(t)<\infty)=\lim_{n\to\infty}\mathsf{P}(N(t)<n)=\lim_{n\to\infty}\mathsf{P}(S_n>t)=1.$$

所以$N(t)$取有限值.

容易看出，更新过程$\{N(t), t\geqslant 0\}$为简单计数过程的充要条件是$\mathsf{P}(W_1=0)=0$，即$F(0)=0$.

对更新过程$N=\{N(t), t\geqslant 0\}$来说，$\{S_n\}$是一列使N"重新开始"的时刻，即对每个S_n，$\{N(t+S_n)-N(S_n), t\geqslant 0\}$与$\{N(t), t\geqslant 0\}$是同分布的. 这一点在研究更新过程时常常用到. S_n的分布函数

$$\mathrm{P}(S_n \leqslant t) = \underbrace{F * \cdots * F}_{n\text{次}}(t) = F^{*n} \tag{5}$$

是 F 的 n 重卷积，$*$ 是卷积的记号.按惯例，F 的“零次”卷积 F^{*0}为

$$F^{*0}(t) = \begin{cases} 1, & t \geqslant 0, \\ 0, & t < 0, \end{cases}$$

它是以概率 1 为零的随机变量的分布函数，由(3)，$N(t)$的分布为

$$\begin{aligned} \mathrm{P}(N(t) = n) &= \mathrm{P}(S_n \leqslant t < S_{n+1}) = \mathrm{P}(S_n \leqslant t) - \mathrm{P}(S_{n+1} \leqslant t) \\ &= F^{*n}(t) - F^{*(n+1)}(t), \qquad n \geqslant 0. \end{aligned}$$

我们早已熟知，若等待时间分布为指数分布，则更新过程就是泊松过程.

一种元件(例如晶体管、灯泡等等)使用若干时间后损坏失效，在失效后随即换上新的同样的元件(假定更换是瞬时间完成的，即不占用时间)，各个元件的寿命是相互独立同分布的，则到时刻 t 为止失效的元件个数，或已发生的更新次数是一个更新过程.这就是更新过程的名称的由来.然而，应用更新过程的场合可以是多种多样的.例如，公路上的车流从距离上看常被考虑为一个更新过程，即把前后车辆间的距离假定为相互独立同分布的随机变量列.另一方面，在一固定地点观察通过该点的累计车辆数，也常假定为是一个更新过程.这相当于假定通过该点的车辆的时间间隔是相互独立同分布的.

例 1.1　Ⅰ型计数器　记数器是一个能够记录接受到的信号次数的装置.例如，盖革(Geiger)－谬勒(Müller)计数器是一个记录接受到的放射性粒子个数的装置.计数器接受到一个信号会使计数器关闭一段时间不能再记录新到的信号.假定在计数器关闭时收到的信号对计数器不起作用，这样的计数器称为Ⅰ型计数器.设$\{W_n, n \geqslant 1\}$为 i.i.d.严格正的随机变量序列.令 $S_0 = 0, S_n = W_1 + \cdots + W_n, n \geqslant 1$. $S_n, n \geqslant 0$ 是第 n 个信号到达的时刻.这里我们假定 $S_0 = 0$ 时就有一个信号到来.设$\{Y_n, n \geqslant 0\}$为 i.i.d.严格正的随机变量序列，且与$\{W_n, n \geqslant 1\}$相互独立. Y_n 是第 n 个被记录下来的信号使计数器关闭的时间.计数器操作如下：在初始时刻记录下一个信号，接着关闭 Y_0 时间.这段时间过后第一个来到的信号又被记录下来，并使计数器关闭 Y_1 时间.之后再继续重复进行.定义

$$n_0 = 0, \quad n_j = \min\{k > n_{j-1} : S_k > S_{n_{j-1}} + Y_{j-1}\}, \quad j \geqslant 1,$$

则 S_{n_j} 是第 j 次记录的时刻.由于从每个记录时刻开始，计数器的操作就像重新开始一样，因此$\{n_j - n_{j-1}, j \geqslant 1\}$与$\{S_{n_j} - S_{n_{j-1}}, j \geqslant 1\}$都是 i.i.d.随机变量序列.(如读者有兴趣证明这事实，可参见第一章的定理 2.2.)令

$$M(t) = \max\{j \geqslant 0 : S_{n_j} \leqslant t\}, \qquad t \geqslant 0,$$

则$\{M(t), t \geqslant 0\}$也是更新过程，$M(t)$是在$(0, t]$中记录下来的信号个数.　□

例 1.2　Ⅱ型计数器　若在计数器关闭时来到的信号也要对计数器起作用，使它关闭一段时间，这样的计数器称为Ⅱ型计数器.引进与上例相同的随机变量，但其中的 Y_n 是第 n 个信号使计数器关闭的时间.现在，对每个 $k \geqslant 0$，$(S_k, S_k + Y_k]$都是计数器关闭的时间，第 k 个信号被

记录下来意味着 $S_k > S_r + Y_r, r = 0,1,\cdots,k-1$. 定义

$$n_0 = 0, n_j = \inf\{k > n_{j-1} : S_k > S_r + Y_r, r = n_{j-1}, \cdots, k-1\}, j \geq 1,$$

则 S_{n_j} 是第 j 次记录的时刻. 由于从每个记录时刻开始,计数器的操作就像重新开始一样,假定 $\mathsf{P}(n_1 < \infty) = 1$,那么 $\{n_j - n_{j-1}, j \geq 1\}$ 与 $\{S_{n_j} - S_{n_{j-1}}, j \geq 1\}$ 都是 i.i.d. 随机变量序列. 仍以 $M(t)$ 记 $(0,t]$ 中记录下来的信号个数,则 $\{M(t), t \geq 0\}$ 也是更新过程. □

例 1.3 排队系统 $G/G/1$ 这时来到排队系统的顾客的累计数就是一个更新过程. 凡有顾客在接受服务就称系统处于忙碌状态,系统中没有顾客就称系统处于空闲状态. 假定在初始时刻就有一个顾客到来,第一个忙碌周期已经开始,然后直到系统中的顾客全部服务完毕,第一个忙碌周期结束,接着第一个空闲周期开始. 但当又有顾客到来时又一个忙碌周期开始,这时系统的状况又重新开始.

引进与例 1.1 相同的随机变量,但其中的 $S_n, n \geq 0$ 是第 n 个顾客来到的时刻, $Y_n, n \geq 0$ 是第 n 个顾客的服务时间. 定义

$$n_0 = 0, n_j = \inf\{k > n_{j-1} : S_k > S_{n_{j-1}} + Y_{n_{j-1}} + \cdots + Y_{k-1}\}, j \geq 1,$$

则 S_{n_j} 是忙碌周期开始的时刻. 由于从每个忙碌周期的开始时刻起,排队系统的状况就像重新开始一样,假定 $\mathsf{P}(n_1 < \infty) = 1$(例如, $\mathsf{E}[W_1] > \mathsf{E}[Y_1]$ 时这可由强大数律得到),那么 $\{n_j - n_{j-1}, j \geq 1\}$ 与 $\{S_{n_j} - S_{n_{j-1}}, j \geq 1\}$ 都是 i.i.d. 随机变量序列. 因此, $S_{n_j}, j \geq 1$, 也构成更新时刻. □

定理 1.1 对更新过程 $N = \{N(t), t \geq 0\}$,以概率 1 有

$$\lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{\mu}, \tag{6}$$

$\mu = \infty$ 时, $1/\mu$ 理解为 0.

证 由 $N(t)$ 的定义可知 $\lim_{t \to \infty} N(t) = \infty$ 及

$$S_{N(t)} \leq t \leq S_{N(t)+1},$$

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)} = \frac{S_{N(t)+1}}{N(t)+1} \cdot \frac{N(t)+1}{N(t)}. \tag{7}$$

由强大数律,以概率 1 有

$$\lim_{t \to \infty} \frac{S_{N(t)}}{N(t)} = \mathsf{E}\, W_1 = \mu.$$

从而由(7)即得(6). □

定理 1.2 设更新过程 $N = \{N(t), t \geq 0\}$ 满足条件:

$$\mathsf{E}[W_1] = \mu < \infty, \qquad \mathsf{Var}[W_1] = \sigma^2 < \infty,$$

则对一切实数 y

$$\lim_{t \to \infty} \mathsf{P}\left(\frac{N(t) - t/\mu}{\sigma\sqrt{t/\mu^3}} \leq y\right) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \mathrm{d}x. \tag{8}$$

证 令 $r_t = t/\mu + y\sigma\sqrt{t/\mu^3}$, $\hat{r}_t = [r_t] + 1$，则

$$\mathrm{P}\left(\frac{N(t) - t/\mu}{\sigma\sqrt{t/\mu^3}} \leqslant y\right) = \mathrm{P}(N(t) \leqslant [r_t]) = \mathrm{P}(N(t) < \hat{r}_t)$$

$$= \mathrm{P}(S_{\hat{r}_t} > t) = \mathrm{P}\left\{\frac{S_{\hat{r}_t} - \hat{r}_t\mu}{\sigma\sqrt{\hat{r}_t}} > \frac{t - \hat{r}_t\mu}{\sigma\sqrt{\hat{r}_t}}\right\}.$$

由于 $\lim_{t\to\infty}\hat{r}_t/r_t = 1$，

$$\lim_{t\to\infty}\frac{t - \hat{r}_t\mu}{\sigma\sqrt{\hat{r}_t}} = \lim_{t\to\infty}\frac{t - r_t\mu}{\sigma\sqrt{r_t}} = \lim_{t\to\infty}\frac{-y\sigma\sqrt{t/\mu}}{\sigma\sqrt{t/\mu + y\sigma\sqrt{t/\mu^3}}} = -y,$$

故由中心极限定理

$$\lim_{t\to\infty}\mathrm{P}\left(\frac{N(t) - t/\mu}{\sigma\sqrt{t/\mu^3}} \leqslant y\right) = 1 - \Phi(-y) = \Phi(y),$$

其中 $\Phi(y)$是正态分布函数. □

定理 1.2 是更新过程的中心极限定理：$N(t)$是渐近正态分布的，其渐近均值为 t/μ，渐近方差为 $t\sigma^2/\mu^3$.

定理 1.3 更新过程$\{N(t), t \geqslant 0\}$的任意阶矩存在：对一切 $t, r > 0$，

$$\mathrm{E}[N(t)]^r < \infty.$$

特别，

$$m(t) = \mathrm{E}[N(t)] = \sum_{n=1}^{\infty} F^{*n}(t) < \infty. \tag{9}$$

证 对固定的 $t > 0$，

$$\mathrm{E}[N(t)]^r \leqslant \sum_{n=1}^{\infty} n^r \mathrm{P}(N(t) \geqslant n) = \sum_{n=1}^{\infty} n^r \mathrm{P}(S_n \leqslant t) = \sum_{n=1}^{\infty} n^r F^{*n}(t).$$

因为 $S_n \to \infty$，故

$$\lim_{n\to\infty}\mathrm{P}(S_n \leqslant t) = 0, \qquad \forall\, t \geqslant 0.$$

因而存在正整数 n 使 $a = F^{*n}(t) < 1$. 又因对一切 $m, n \geqslant 0$ 有

$$F^{*(m+n)}(t) = \int_0^t F^{*m}(t-s)\mathrm{d}F^{*n}(s) \leqslant F^{*m}(t)F^{*n}(t), \tag{10}$$

$$F^{*(lm+k)}(t) \leqslant a^l, \qquad k = 0, 1, \cdots, m-1.$$

这时

$$\mathrm{E}[N(t)]^r \leqslant \sum_{n=1}^{\infty} n^r F^{*n}(t) = \sum_{l=0}^{\infty}\sum_{k=0}^{m-1}(lm+k)^r F^{*(lm+k)}(t)$$

$$\leqslant m^r \sum_{l=0}^{\infty}(l+1)^r a^l < \infty.$$

同时

$$m(t)=\mathrm{E}[N(T)]=\sum_{n=1}^{\infty}\mathrm{E}[1_{\{S_n\leqslant t\}}]=\sum_{n=1}^{\infty}\mathrm{P}(S_n\leqslant t)=\sum_{n=1}^{\infty}F^{*n}(t). \qquad \square$$

定义 $m(t)=\mathrm{E}[N(t)]$称为更新过程 $N=\{N(t),t\geqslant 0\}$的更新函数.

$m(t)$是$[0,\infty)$上的非负右连续单调增加函数(事实上,级数 $\sum_{n=1}^{\infty}F^{*n}(t)$在有限区间上一致收敛),且

$$\begin{aligned} m(0)&=\sum_{n=1}^{\infty}F^{*n}(0)=\sum_{n=1}^{\infty}\mathrm{P}(S_n=0)\\ &=\sum_{n=1}^{\infty}\mathrm{P}(W_1=0,\cdots,W_n=0)\\ &=\sum_{n=1}^{\infty}[F(0)]^n=\frac{F(0)}{1-F(0)}. \end{aligned} \tag{11}$$

在上面涉及的分布函数的卷积((5)及(10)式)和本章下面的讨论中,我们都要用到斯蒂尔切斯(Stieltjes)积分:$\int_0^t h(s)\mathrm{d}F(s),0\leqslant t\leqslant\infty$.对于不熟悉斯蒂尔切斯积分的读者,可像学习初等概率论时那样,先将分布函数 F 分为连续型及离散型分别讨论.若 F 有分布密度f,则

$$\int_0^t h(s)\mathrm{d}F(s)=\int_0^t h(s)f(s)\mathrm{d}s$$

归结为普通的积分.若 F 是离散型分布:

$$F(t)=\sum_{j:t_j\leqslant t}p_j,\quad t_j\geqslant 0,p_j>0,j\geqslant 1,\quad \sum_j p_j=1,$$

则

$$\int_0^t h(s)\mathrm{d}F(s)=\sum_{j:t_j\leqslant t}h(t_j)p_j.$$

这时需注意,上式中的积分区域是$[0,t]$,包含了零点.因此,当 $t=0$ 时,这积分等于 $h(0)F(0)$.但当 $0<a<b$ 时,为了使积分有可加性,应定义

$$\int_a^b h(s)\mathrm{d}F(s)=\sum_{j:a<t_j\leqslant b}h(t_j)p_j,$$

即积分区域是$(a,b]$.一般地,若 $F=\lambda F_1+(1-\lambda)F_2$,$0<\lambda<1$,其中 F_1 及 F_2 分别为连续型及离散型分布函数,则

$$\int_0^t h(s)\mathrm{d}F(s)=\lambda\int_0^t h(s)\mathrm{d}F_1(s)+(1-\lambda)\int_0^t h(s)\mathrm{d}F_2(s).$$

今后我们还要用到斯蒂尔切斯积分:$\int_0^t h(s)\mathrm{d}m(s)$.定义完全是类似的.由于 $\mathrm{d}F^{*n}(t)$是在$(t,t+\mathrm{d}t]$中发生第 n 次更新的概率,因此 $\mathrm{d}m(t)=\sum_{n=1}^{\infty}\mathrm{d}F^{*n}(t)$是在$(t,t+\mathrm{d}t]$中发生更新的概率.这样,$\mathrm{d}m(t)$有鲜明的概率意义.

定义 $$\phi(\lambda)=\mathrm{E}[\mathrm{e}^{-\lambda W_1}]=\int_0^{\infty}\mathrm{e}^{-\lambda t}\mathrm{d}F(s),\quad \lambda>0$$

称为 F 或 W_1 的拉普拉斯变换.若 F 有密度函数 f,则上面定义的就是 f 的拉普拉斯变换(参见第三章第二节).由于 $F(0)<1$,

$$0<\phi(\lambda)<1,\qquad \lambda>0.$$

引理 1.1 分布函数 F 为其拉普拉斯变换 $\phi(\lambda),\lambda>0$ 唯一决定.

证 定义

$$G(t)=\begin{cases}0, & t\leqslant 0,\\ 1-F((-\ln t)-0), & 0<t\leqslant 1,\\ 1, & 1<t,\end{cases}$$

则 G 为(0,1]上的概率分布函数(注意,为了使 G 右连续,在上面的定义式中用了左极限),且

$$\phi(\lambda)=\int_0^1 t^{\lambda}\mathrm{d}G(t),\qquad \lambda>0.$$

G 的特征函数 $g(u)=\int_0^1\mathrm{e}^{\mathrm{i}ut}\mathrm{d}G(t)$ 在 $u=0$ 点的 n 阶导数即为 $\mathrm{i}^n\phi(n)$,而 $\{\phi(n),n\geqslant 1\}$ 有界,因此

$$g(u)=1+\sum_{n=1}^{\infty}\frac{\phi(n)}{n!}(\mathrm{i}u)^n,\qquad -\infty<u<\infty.$$

所以 $g(u)$ 为 $\{\phi(n),n\geqslant 1\}$ 所唯一决定,即 G 为 ϕ 所唯一决定,从而

$$F(t)=1-G((\mathrm{e}^{-t})-0),\qquad t\geqslant 0$$

为 $\phi(\lambda),\lambda>0$,唯一决定. □

定理 1.4 等待时间分布函数 F 由更新函数 m 唯一确定.

证 由于

$$\int_0^{\infty}\mathrm{e}^{-\lambda t}\mathrm{d}F^{*n}(t)=\mathrm{E}\,\mathrm{e}^{-\lambda S_n}=\prod_{i=1}^{n}\mathrm{E}\,\mathrm{e}^{-\lambda W_i}=[\phi(\lambda)]^n,\quad \lambda,n\geqslant 1,$$

因此 $m(t)$ 的拉普拉斯变换为

$$\widetilde{m}(\lambda)=\int_0^{\infty}\mathrm{e}^{-\lambda t}\mathrm{d}m(t)=\sum_{n=1}^{\infty}\int_0^{\infty}\mathrm{e}^{-\lambda t}\mathrm{d}F^{*n}(t)=\sum_{n=1}^{\infty}[\phi(\lambda)]^n=\frac{\phi(\lambda)}{1-\phi(\lambda)}.$$

所以

$$\phi(\lambda)=\frac{\widetilde{m}(\lambda)}{1+\widetilde{m}(\lambda)}$$

由 m 唯一确定,F 亦随之唯一确定. □

推论 若更新过程 N 的更新函数 $m(t)=\lambda t$,则 N 是参数为 λ 的泊松过程.

证 若 N 为泊松过程,则 $N(t)$ 服从参数为 λt 的泊松分布,从而 $m(t)=\lambda t$.反之,若更新

过程 N 的更新函数 $m(t)=\lambda t$,则由定理 1.4,等待时间分布 F 是参数为 λ 的指数分布,从而 N 为泊松过程. □

定理 1.5 更新函数 $m(t)$满足下列方程:

$$m(t)=F(t)+\int_0^t m(t-y)\mathrm{d}F(y). \tag{12}$$

证

$$\begin{aligned} m(t) &= \sum_{n=1}^{\infty} F^{*n}(t)=F(t)+\sum_{n=2}^{\infty} F^{*(n-1)} * F(t) \\ &= F(t)+m * F(t)=F(t)+\int_0^t m(t-y)\mathrm{d}F(y). \end{aligned}$$ □

我们还要利用更新过程的特点,给出定理 1.5 的第二种证法,为此需要用到一般形式的全概率(或全期望)公式,我们将它写成一条引理.

引理 1.2 设 X,Y 为随机变量,X 的分布函数为 G,则

$$\mathrm{E}[Y]=\int \mathrm{E}[Y\mid X=x]\mathrm{d}G(x). \tag{13}$$

设 A 为随机事件,在(13)式中取 $Y=1_A$ 即得

$$\mathrm{P}(A)=\int \mathrm{P}[A\mid X=x]\mathrm{d}G(x). \tag{14}$$

若 X 为离散型随机变量,(14)式就归结为通常的全概率公式.(13)式的证明还有赖于如何定义其中的条件期望.下面作为参考,我们只就(X,Y)为连续型变量的情形,按通常条件期望的定义,给出(13)式的论证.

设(X,Y)有联合密度函数 $k(x,y)$.这时 X 有密度函数 $g(x)=\int k(x,y)\mathrm{d}y$,$Y$ 有密度函数 $h(y)=\int k(x,y)\mathrm{d}x$,从而

$$\begin{aligned} \mathrm{E}[Y\mid X=x] &= \int y\frac{k(x,y)}{g(x)}\mathrm{d}y, \\ \int \mathrm{E}[Y\mid X=x]\mathrm{d}G(x) &= \int\left(\int y\frac{k(x,y)}{g(x)}\mathrm{d}y\right)g(x)\mathrm{d}x \\ &= \int\left(\int yk(x,y)\mathrm{d}y\right)\mathrm{d}x=\int y\left(\int k(x,y)\mathrm{d}x\right)\mathrm{d}y \\ &= \int yh(y)\mathrm{d}y=\mathrm{E}[Y]. \end{aligned}$$

在实际上,使用(13)或(14)式往往是因为根据具体情况,条件期望 $\mathrm{E}[Y\mid X=x]$或条件概率 $\mathrm{P}[A\mid X=x]$可直接先得到,并非像上面的论证中那样要借助于联合分布.这在下面的讨论中将看得很清楚.

现在,我们再回到定理 1.5 的证明上来.

$$m(t)=\mathrm{E}[N(t)]=\int_0^{\infty} \mathrm{E}[N(t)\mid W_1=s]\mathrm{d}F(s). \tag{15}$$

注意到，在 $W_1=s>t$ 时，显然有 $N(t)=0$，因此 $\mathrm{E}[N(t)|W_1=s]=0$. 在 $W_1=s\leqslant t$ 时，以时刻 s 作为新的时间起点，那么更新过程就可以看作是重新开始，计入第一个更新，就得到 $\mathrm{E}[N(t)|W_1=s]=1+m(t-s)$. 将所得结果代入(15)式即得(12)式：

$$m(t)=\int_0^t[1+m(t-s)]\mathrm{d}F(s)=F(t)+\int_0^t m(t-s)\mathrm{d}F(s).$$

上述证明方法是一种概率证法，其关键步骤是对第一个更新时刻 W_1 取条件期望. 这个方法充分利用了更新过程的构造特点，即从一个更新时刻起，在概率上更新过程好像重新开始一样. 这个方法通常被称为*更新技巧*，是研究更新过程的主要方法之一，今后将被频繁地使用.

定义 设 $b(t)$ 为 $[0,\infty)$ 上的局部有界函数（在任一有界区间上有界的函数），未知函数 $B(t)$ 的下列方程：

$$B(t)=b(t)+\int_0^t B(t-s)\mathrm{d}F(s) \tag{16}$$

称为*更新方程*.

更新过程的更新函数满足的方程(12)就是一个更新方程. 使用更新技巧还将导出各种更新方程. 因此，我们先讨论更新方程的存在唯一性.

定义 设 $h(t)$ 为 $[0,\infty)$ 上的局部有界函数，$G(t)$ 为 $[0,\infty)$ 上的右连续单调增加函数. 令

$$h*G(t)=\int_0^t h(t-s)\mathrm{d}G(s),\qquad t\geqslant 0.$$

注意，按定义 $h*G(0)=h(0)G(0)$. 易见 $h*G$ 仍为局部有界函数. 当 h 是右连续单调增加函数时，$h*G$ 也是右连续单调增加函数.

引理 1.3 设 $h(t)$ 为 $[0,\infty)$ 上的局部有界函数，$G(t)$，$H(t)$ 为 $[0,\infty)$ 上的右连续单调增加函数，则

(1) $G*H=H*G$，

(2) $(h*G)*H=h*(G*H)$.

证 不失一般性，可假设 G 与 H 是 $[0,\infty)$ 上的概率分布函数，且它们分别是独立随机变量 X 与 Y 的分布函数. 这时 $G*H$ 是 $X+Y$ 的分布函数，因此与 $H*G$ 相同. 另一方面，

$$\begin{aligned}(h*G)*H(t)&=\int_0^t\left(\int_0^{t-s}h(t-s-u)\mathrm{d}G(u)\right)\mathrm{d}H(s)\\&=\int_0^t\int_0^t h(t-(s+u))1_{\{s+u\leqslant t\}}\mathrm{d}G(u)\mathrm{d}H(s)\\&=\mathrm{E}[h(t-(X+Y))1_{\{X+Y\leqslant t\}}]\\&=\int_0^t h(t-s)\mathrm{d}(G*H)(s)\\&=h*(G*H)(t).\qquad\square\end{aligned}$$

定理 1.6 设 $b(t)$ 为局部有界函数，则更新方程(16)有唯一的局部有界解

$$B(t)=b(t)+\int_0^t b(t-s)\mathrm{d}m(s). \tag{17}$$

证 已经知道(17)式定义的 B 是局部有界的,且

$$\begin{aligned} B &= b + b * m = b * F + b * \sum_{n=2}^{\infty} F^{*n} \\ &= b + \left(b + \sum_{n=1}^{\infty} b * F^{*n}\right) * F = b + (b + b * m) * F \\ &= b + B * F. \end{aligned}$$

下证唯一性.对满足(16)的局部有界的 B,

$$\begin{aligned} B &= b + B * F = b + b * F + B * F * F = \cdots \\ &= b + b * \sum_{j=1}^{n-1} F^{*j} + B * F^{*n}, \end{aligned} \tag{18}$$

$$\sup_{0 \leqslant s \leqslant t} |B * F^{*n}(s)| \leqslant \sup_{0 \leqslant s \leqslant t} |B(s)| F^{*n}(t), \qquad \forall t > 0.$$

由于 $S_n \to \infty$,

$$\lim_{n \to \infty} F^{*n}(t) = \mathrm{P}(S_n \leqslant t) = 0,$$

故

$$\lim_{n \to \infty} B * F^{*n}(t) = 0.$$

再在(18)中令 $n \to \infty$ 得

$$B = b + b * \sum_{j=1}^{\infty} F^{*j} = b + b * m,$$

即 B 必有(17)的形式. □

推论 对更新过程 $\{N(t), t \geqslant 0\}$,

$$\mathrm{E}[N(t)]^2 = m(t) + 2\int_0^t m(t-s)\mathrm{d}m(s), \quad t \geqslant 0. \tag{19}$$

证 记 $B(t) = \mathrm{E}[N(t)]^2$,则

$$B(t) = 2\sum_{n=1}^{\infty} n\mathrm{P}(N(t) \geqslant n) - \mathrm{E}\,N(t) = \sum_{n=1}^{\infty} (2n-1)F^{*n}(t),$$

$$\begin{aligned} B * F &= \sum_{n=1}^{\infty} (2n-1)F^{*n} * F = \sum_{n=1}^{\infty} (2n-1)F^{*(n+1)} \\ &= \sum_{n=2}^{\infty} [(2n-1)-2]F^{*n} \\ &= \sum_{n=1}^{\infty} (2n-1)f^{*n} - 2\sum_{n=1}^{\infty} F^{*n} + F \\ &= B - 2m + F, \end{aligned}$$

即 B 满足更新方程

$$B = (2m - F) + B * F,$$

因而

$$\begin{aligned} B &= (2m - F) + (2m - F) * m \\ &= 2m + 2m * m - (F + F * m) \\ &= 2m + 2m * m - m = m + 2m * m, \end{aligned}$$

即得(19),这里利用了 $m = F + F * m$,因为 m 满足更新方程 $m = F + m * F$. □

定理 1.7 设 $F(t)$有局部有界的密度函数 $f(t)$:

$$F(t) = \int_0^t f(s)\mathrm{d}s, \qquad t \geqslant 0,$$

则有

$$m(t) = \int_0^t u(s)\mathrm{d}s, \qquad t \geqslant 0, \tag{20}$$

其中

$$u(t) = f(t) + \int_0^t f(t-s)\mathrm{d}m(s), \qquad t \geqslant 0. \tag{21}$$

证 (21)式定义的 $u(t)$是更新方程

$$B(t) = f(t) + \int_0^t B(t-s)\mathrm{d}F(s)$$

的局部有界解,所以有

$$\begin{aligned} \int_0^t u(s)\mathrm{d}s &= \int_0^t f(s)\mathrm{d}s + \int_0^t \mathrm{d}s \int_0^s u(s-r)\mathrm{d}F(r) \\ &= F(t) + \int_0^t \left(\int_r^t u(s-r)\mathrm{d}s\right)\mathrm{d}F(r) \\ &= F(t) + \int_0^t \left(\int_0^{t-r} u(s)\mathrm{d}s\right)\mathrm{d}F(r). \end{aligned}$$

这说明 $\int_0^t u(s)\mathrm{d}s$ 是更新方程 $B = F + B * F$ 的解,但 $\int_0^t u(s)\mathrm{d}s$ 是局部有界的,所以由定理 1.5 及更新方程解的唯一性有

$$m(t) = \int_0^t u(s)\mathrm{d}s. \qquad \square$$

习 题

1-1. 设$\{\xi_n, n \geqslant 1\}$为伯努利过程,

$$N(t) = \sum_{j=1}^{[t]} \xi_j, \qquad t \geqslant 0,$$

则$\{N(t), t \geqslant 0\}$是更新过程,其等待时间分布为几何分布.

1 - 2. 设更新过程$\{N(t),t\geqslant 0\}$的等待时间分布是泊松分布,即:

$$\mathrm{P}(X_n=k)=\frac{\mu^k}{k!}\mathrm{e}^{-\mu},\qquad k=0,1,2,\cdots.$$

求 S_n 的分布,并计算 $\mathrm{P}(N(t)=n)$.

1 - 3. 设更新过程$\{N(t),t\geqslant 0\}$的等待时间有密度函数

$$f(x)=\begin{cases}\rho\mathrm{e}^{-\rho(x-\delta)}, & x>\delta,\\ 0, & x\leqslant\delta,\end{cases}$$

其中 $\delta>0,\rho>0$. 计算 $\mathrm{P}(N(t)\geqslant k)$.

1 - 4. 对更新过程$\{N(t),t\geqslant 0\}$证明:若 $0<\gamma<1/F(0)$,则

$$\sum_{n=1}^{\infty}\gamma^n F^{*n}(t)<\infty,\quad t\geqslant 0.$$

1 - 5. 设更新过程$\{N(t),t\geqslant 0\}$的等待时间有密度函数 $f(x)=\lambda^2 x\mathrm{e}^{-\lambda x},x\geqslant 0$. 试证相应的更新函数为

$$m(t)=\mathrm{E}[N(t)]=\frac{\lambda t}{2}-\frac{1}{4}(1-\mathrm{e}^{-2\lambda t}).$$

1 - 6. 证明:更新过程的更新函数 $m(t)$连续的充要条件是它的等待时间分布函数 $F(t)$连续.

1 - 7. 证明:更新过程 $N=\{N(t),t\geqslant 0\}$的更新函数 $m(t)$满足:

(1) $\dfrac{1}{\mathrm{E}[\min(W_1,t)]}\leqslant\dfrac{1+m(t)}{t}\leqslant\dfrac{2}{\mathrm{E}[\min(W_1,t)]},\quad t>0;$

(2) $m(t)-m(t-h)\leqslant 1+m(h),0\leqslant h\leqslant t.$

1 - 8. 对更新过程$\{N(t),t\geqslant 0\}$,令 $m_k(t)=\mathrm{E}[N(t)]^k,k\geqslant 1$. 证明:对 $k\geqslant 2,m_k(t)$是下列更新方程的解:

$$z(t)=h_k(t)+\int_0^t z(t-s)\mathrm{d}F(s),$$

其中 $h_k(t)=(-1)^{k-1}[F(t)-\mathrm{C}_k^1 m_1(t)+\cdots+(-1)^{k-1}\mathrm{C}_k^{k-1}m_{k-1}(t)]$.

1 - 9. 设 $b(t)$为局部有界函数,$B(t)$是更新方程

$$B(t)=b(t)+\int_0^t B(t-s)\mathrm{d}F(s)$$

的解,则 $Z(t)=\int_0^t B(s)\mathrm{d}s$ 是更新方程

$$Z(t)=\int_0^t b(s)\mathrm{d}s+\int_0^t Z(t-s)\mathrm{d}F(s)$$

的解.

1 - 10. 设在Ⅱ型计数器中,信号来到的间隔服从参数为 λ 的指数分布. 每个来到的信号使

计数器关闭一段固定长度 d 的时间. 记 $B(t)$ 为时刻 t 计数器不关闭的概率. 证明:

$$B(t)=\begin{cases}0, & t\leqslant d,\\ e^{-\lambda d}, & t>d.\end{cases}$$

1-11. 设 N_1 与 N_2 为两个独立的更新过程, 有相同的连续等待时间分布函数 F. 若 $N=N_1+N_2$ 也为更新过程, 则 N_1 及 N_2 从而 N 都是泊松过程.

1-12. 设 N_1 为泊松过程, N_2 为更新过程, N_1 与 N_2 相互独立. 若 $N=N_1+N_2$ 也为更新过程, 则 N_2 从而 N 都是泊松过程.

§4.2 延迟更新过程

在本节中我们将更新过程略作推广, 讨论延迟更新过程.

定义 设 $W=\{W_j, j\geqslant 1\}$ 为一列独立的非负随机变量列, $\{W_j, j\geqslant 2\}$ 为同分布的, 分布为 F, W_1 的分布为 G. 仍令 $S_0=0, S_n=\sum_{j=1}^{n} W_j, n\geqslant 1$,

$$N(t)=\sup\{n: S_n\leqslant t\},$$

则 $N=\{N(t), t\geqslant 0\}$ 称为延迟更新过程, $\{S_n, n\geqslant 1\}$ 仍称为更新时刻, $\{W_j, j\geqslant 1\}$ 为更新(跳跃)间隔或等待时间, F 为等待时间分布. 当 $F=G$ 时, 延迟更新过程就是更新过程.

延迟更新过程最初来源于这样的考虑: 在观察更新元件的初始时刻, 第一个元件已经使用了 t_0 时间, 因此第一个等待时间的分布应修正为

$$G(t)=\frac{F(t+t_0)-F(t_0)}{1-F(t_0)}.$$

只要 F 不是指数分布, 它与 F 已不相同了, 这就是"延迟"的原意. 我们定义的延迟更新过程更为一般, 适用于更广泛的场合.

延迟更新过程的结构与更新过程并无差别, 故它仍为计数过程, 仍成立关系式

$$\{N(t)\geqslant n\}=\{S_n\leqslant t\}, \qquad \{N(t)=n\}=\{S_n\leqslant t<S_{n+1}\},$$

等等. 事实上, 延迟更新过程在第一次更新之后与更新过程就没有区别了. 所以, 可以预料更新过程的许多结果对于延迟更新过程也成立, 也可以用同样的方法(如更新技巧)讨论延迟更新过程.

定理 2.1 设 $N=\{N(t), t\geqslant 0\}$ 为延迟更新过程, 则对一切 $t\geqslant 0$, $m(t)=\mathrm{E}[N(t)]<\infty$, 且 $m(t)$ 是更新方程

$$m(t)=G(t)+\int_0^t m(t-s)\mathrm{d}F(s) \tag{1}$$

的唯一局部有界解:

$$m=G+G*m_F=\sum_{n=1}^{\infty} G*F^{*(n-1)}, \tag{2}$$

其中

$$m_F(t) = \sum_{n=1}^{\infty} F^{*n}(t). \tag{3}$$

证 对延迟更新过程我们仍有

$$\begin{aligned} m(t) = \mathrm{E}[N(t)] &= \sum_{n=1}^{\infty} \mathrm{P}(N(t) \geqslant n) = \sum_{n=1}^{\infty} \mathrm{P}(S_n \leqslant t) \\ &= G(t) + \sum_{n=1}^{\infty} G * F^{*n}(t) = G(t) + \int_0^t G(t-s)\mathrm{d}m(s). \end{aligned}$$

这里只要注意，$G * F^{*n}(t) \leqslant F^{*n}(t), t \geqslant 0$,

$$m(t) = \sum_{n=0}^{\infty} G * F^{*n} \leqslant \sum_{n=0}^{\infty} F^{*n}(t) = 1 + m_F(t) < \infty, \quad t \geqslant 0.$$

由此即知 $m(t)$是更新方程(1)的唯一局部有界解. □

对延迟更新过程的等待时间序列$\{W_n, n \geqslant 1\}$，强大数律与中心极限定理仍然适用，因此定理 1.1 与 1.2 以及它们的证明对延迟更新过程依然成立.

定理 2.2 设 $N = \{N(t), t \geqslant 0\}$为延迟更新过程，则以概率 1 有

$$\lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{\mu},$$

其中 $\mu = \mathrm{E}[W_2]$.

定理 2.3 设延迟更新过程 $N = \{N(t), t \geqslant 0\}$满足条件

$$\mu = \mathrm{E}[W_2] < \infty, \qquad \sigma^2 = \mathrm{Var}[W_2] < \infty,$$

则对一切实数 y,

$$\lim_{t \to \infty} \mathrm{P}\left(\frac{N(t) - t/\mu}{\sigma \sqrt{t/\mu^3}} \leqslant y \right) = \int_0^y \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \mathrm{d}x.$$

定义 对延迟更新过程 $N = \{N(t), t \geqslant 0\}$,

$$\begin{aligned} &A(t) = t - S_{N(t)}, \\ &R(t) = S_{N(t)+1} - t, \\ &\beta(t) = A(t) + R(t) = S_{N(t)+1} - S_{N(t)} \end{aligned}$$

分别称为时刻 t 的年龄，剩余寿命和总寿命. 它们分别是时刻 t 正在使用中的元件的年龄，剩余寿命和总寿命. 显然有

$$0 \leqslant A(t) \leqslant t, \qquad t \geqslant 0. \tag{4}$$

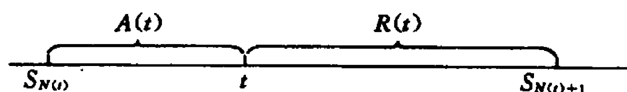


图 3

年龄过程与剩余寿命过程间有下列重要的关系:对任意的 $x\geqslant 0, 0\leqslant y\leqslant t$,

$$\begin{aligned}\{R(t)>x, A(t)\geqslant y\} &= \{(t-y, t+x]\text{中 } N \text{ 无更新}\} \\ &= \{R(t-y)>x+y\} = \{A(t+x)\geqslant x+y\}.\end{aligned} \tag{5}$$

在(5)式中取 $y=0$ 得

$$\{R(t)>x\} = \{A(t+x)\geqslant x\}, \qquad x\geqslant 0. \tag{6}$$

下面我们将推导 $A(t), R(t), \beta(t)$ 的分布及期望的公式.

今后,对任一$[0,\infty)$上的分布函数 H,记

$$\overline{H}(t) = 1-H(t), \qquad t\geqslant 0. \tag{7}$$

$$\mu_H = \int_0^\infty \overline{H}(t)\mathrm{d}t. \tag{8}$$

定理 2.4 对延迟更新过程 N,

$$\mathsf{P}(A(t)\geqslant x) = \overline{G}(t) + \int_0^{t-x} \overline{F}(t-s)\mathrm{d}m(s), \qquad t\geqslant x\geqslant 0. \tag{9}$$

证 先设 N 为更新过程. 令 $B(t)=\mathsf{P}(A(t)\geqslant y)$. 用更新技巧

$$B(t) = \int_0^\infty \mathsf{P}(A(t)\geqslant y \mid W_1 = s)\mathrm{d}F(s),$$

$$\mathsf{P}(A(t)\geqslant y\mid W_1=s) = \begin{cases} 1_{\{t\geqslant y\}}, & s>t, \\ \mathsf{P}(A(t-s)\geqslant y), & s\leqslant t. \end{cases}$$

因为 $s>t, y>t$ 时, $N(t)=0, A(t)=t<y$; $s>t, y\leqslant t$ 时, $N(t)=0, A(t)=t\geqslant y$; $s\leqslant t$ 时,将时间原点移至 s,过程重新开始. 所以,

$$B(t) = \overline{F}(t)1_{\{t\geqslant y\}} + \int_0^t B(t-u)\mathrm{d}F(u).$$

这是一个更新方程,由定理 1.6 即得

$$\begin{aligned} B(t) &= \overline{F}(t)1_{\{t\geqslant y\}} + \int_0^t \overline{F}(t-u)1_{\{t-u\geqslant y\}}\mathrm{d}m(u) \\ &= \overline{F}(t)1_{\{t\geqslant y\}} + \int_0^{(t-y)^+} \overline{F}(t-u)\mathrm{d}m(u), \end{aligned}$$

即(9)式成立.

现在讨论一般的延迟更新过程. 仍对第一个更新时刻 W_1 取条件期望,只是现在 W_1 的分布是 G. 对 $x\geqslant 0$

$$\mathsf{P}(A(t)\geqslant x) = \int_0^\infty \mathsf{P}(A(t)\geqslant x\mid W_1=s)\mathrm{d}G(s),$$

$$\mathsf{P}(A(t)\geqslant x\mid W_1=s) = \begin{cases} 1_{\{x\leqslant t\}}, & s>t, \\ g(t-s), & s\leqslant t, \end{cases}$$

利用前面的结果,其中的

$$g=h+h*m_F,\qquad h(t)=\bar{F}(t)1_{\{x\leqslant t\}}.$$

因为 $s>t$ 时,$N(t)=0,A(t)=t$;$s\leqslant t$ 时,以 s 为时间原点,重新开始的过程是一个以 F 为等待时间分布的更新过程.注意到 $g*G=h*(G+m_F*G)=h*m$,所以

$$\mathsf{P}(A(t)\geqslant x)=\bar{G}(t)1_{\{x\leqslant t\}}+\int_0^t\bar{F}(t-s)1_{\{x\leqslant t-s\}}\mathrm{d}m(s).$$

由此即得(9)式. □

推论 更新过程是泊松过程的充要条件为对一切 $t\geqslant 0$,$A(t)$的分布函数是

$$\mathsf{P}(A(t)\leqslant x)\begin{cases}F(x), & x<t,\\ 1, & x\geqslant t,\end{cases}$$

即 $\min\{W_1,t\}$的分布函数.

证 必要性早在第一章的定理 3.6 中证明了.要证充分性.这时 $B(t)=\mathsf{P}(A(t)\geqslant x)=[1-F(x^-)]1_{\{t\geqslant x\}}$.由(9),

$$\begin{aligned}B(t)&=\bar{F}(t)1_{\{t\geqslant x\}}+\int_0^{(t-x)^+}\bar{F}(t-s)\mathrm{d}m(s)\\&=\bar{F}(t)1_{\{t\geqslant x\}}+\int_0^t\bar{F}(t-s)1_{\{t-s\geqslant x\}}\mathrm{d}m(s).\end{aligned}$$

所以

$$B(t)=\bar{F}(t)1_{\{t\geqslant x\}}+\int_0^t B(t-s)\mathrm{d}F(s).$$

因此,$t\geqslant x$ 时

$$\begin{aligned}1-F(x^-)&=\bar{F}(t)+\int_0^t[1-F(x^-)]1_{\{t-s\geqslant x\}}\mathrm{d}F(s)\\&=\bar{F}(t)+[1-F(x^-)]F(t-x),\\&[1-F(x^-)]\bar{F}(t-x)=\bar{F}(t).\end{aligned}$$

所以,$x\geqslant 0,y\geqslant 0$ 时

$$\bar{F}(x+y)=[1-F(x^-)]\bar{F}(y).$$

在上式中取右极限得

$$\bar{F}(x+y)=\bar{F}(x)\bar{F}(y),\qquad x,y\geqslant 0.$$

由此可知,F 是指数分布. □

定理 2.5 对延迟更新过程 N,

$$\mathsf{P}(R(t)>x,A(t)\geqslant y)=\bar{G}(t+x)+\int_0^{t-y}\bar{F}(t+x-s)\mathrm{d}m(s),\quad t,x\geqslant 0,0\leqslant y\leqslant t,\ (10)$$

特别,

$$\mathsf{P}(R(t)>x)=\bar{G}(t+x)+\int_0^t \bar{F}(t-s+x)\mathrm{d}m(s),\qquad t,x\geqslant 0. \tag{11}$$

证 利用(5)及(9)式

$$\begin{aligned}\mathsf{P}(R(t)>x,A(t)\geqslant y)&=\bar{G}(t+x)1_{\{x+y\leqslant t+x\}}+\int_0^{t+x}\bar{F}(t+x-s)1_{\{x+y\leqslant t+x-s\}}\mathrm{d}m(s)\\&=\bar{G}(t+x)1_{\{y\leqslant t\}}+\int_0^{t+x}\bar{F}(t+x-s)1_{\{y\leqslant t-s\}}\mathrm{d}m(s).\end{aligned}$$

由此即得(10)式.在(10)式中令 $y=0$ 即得(11)式. □

推论 更新过程是泊松过程的充要条件为对一切 $t\geqslant 0$,$R(t)$的分布函数就是等待时间分布函数 F.

证 必要性早在第一章的定理3.6中已得到了.要证充分性.这时 $B(t)=\mathsf{P}(R(t)>x)=\bar{F}(x)$.由(11),

$$B(t)=\bar{F}(t+x)+\int_0^t \bar{F}(t+x-s)\mathrm{d}m(s),$$

所以

$$B(t)=\bar{F}(t+x)+\int_0^t B(t-s)\mathrm{d}F(s),$$

即

$$\bar{F}(x)=\bar{F}(t+x)+\int_0^t \bar{F}(x)\mathrm{d}F(s)=\bar{F}(t+x)+\bar{F}(x)F(t),$$

$$\bar{F}(t+x)=\bar{F}(t)\bar{F}(x),\qquad t,x>0,$$

由此即知,F 是指数分布. □

定理 2.6 对延迟更新过程 N,

$$\mathsf{P}(\beta(t)>x)=\bar{F}(\max(t,x))+\int_0^t \bar{F}(\max((t-y),x))\mathrm{d}m(y),\quad t,x\geqslant 0. \tag{12}$$

证 先设 N 为更新过程.令 $B(t)=\mathsf{P}(\beta(t)>x)$,$x\geqslant 0$,仍使用更新技巧.

$$B(t)=\int_0^\infty \mathsf{P}(\beta(t)>x\mid W_1=s)\mathrm{d}F(s).$$

这时

$$\mathsf{P}(\beta(t)>x\mid W_1=s)=\begin{cases}1_{\{s>x\}}, & s>t,\\ \mathsf{P}(\beta(t-s)>x), & s\leqslant t.\end{cases}$$

因为 $s>\max(t,x)$时 $N(t)=0,\beta(t)=s>x$;$\max(t,x)\geqslant s>t$ 时 $N(t)=0,\beta(t)=s\leqslant x$;$t\geqslant s$ 时,将时间原点移至 s,过程重新开始.所以

$$B(t)=\overline{G}(\max(t,x))+\int_0^t B(t-s)\mathrm{d}F(s).$$

由定理 1.6 即得(12)式.

现在讨论一般的延迟更新过程.与定理 2.4 的证明相仿,仍对第一个更新时刻 W_1 取条件期望,而 W_1 的分布是 G.

$$B(t)=\int_0^{\infty}\mathsf{P}(\beta(t)>x\mid W_1=s)\mathrm{d}G(s).$$

这时

$$\mathsf{P}(\beta(t)>x\mid W_1=s)=\begin{cases}1_{\{s>x\}}, & s>t,\\ g(t-s), & s\leqslant t,\end{cases}$$

其中的 $g=h+h*m_F, h(t)=\overline{F}(\max(t,x))$.仍由 $g*G=h*(G+m_F*G)=h*m$ 可得(12)式. □

定理 2.7 设 N 为延迟更新过程,则

$$\mathsf{E}[A(t)]=t\overline{G}(t)+\int_0^t(t-s)\overline{F}(t-s)\mathrm{d}m(s), t\geqslant 0 \tag{13}$$

证 记 $B(t)=\mathsf{E}A(t)$.先设 N 为更新过程.则因

$$\mathsf{E}[A(t)\mid W_1=s]=\begin{cases}t, & s>t,\\ \mathsf{E}[A(t-s)], & s\leqslant t,\end{cases}$$

所以有

$$B(t)=\int_t^{\infty}t\mathrm{d}F(s)+\int_0^t B(t-s)\mathrm{d}F(s)=t\overline{F}(t)+\int_0^t B(t-s)\mathrm{d}F(s).$$

再由定理 1.6 可解得

$$B(t)=t\overline{F}(t)+\int_0^t(t-s)\overline{F}(t-s)\mathrm{d}m(s),$$

即(13)成立.

现在讨论一般的延迟更新过程.与定理 2.4 的证明相仿,

$$B(t)=\int_0^{\infty}\mathsf{E}[A(t)\mid W_1=s]\mathrm{d}G(s).$$

这时

$$\mathsf{E}[A(t)\mid W_1=s)=\begin{cases}t, & s>t,\\ g(t-s), & s\leqslant t,\end{cases}$$

其中的 $g=h+h*m_F, h(t)=\overline{F}(t)$.仍由 $g*G=h*(G+m_F*G)=h*m$ 可得(13)式.
□

定理 2.8 设延迟更新过程 N 满足条件:$\mu_G=\mathsf{E}W_1<\infty, \mu_F=\mathsf{E}[W_2]<\infty$,则

$$\mathsf{E}[R(t)]=\int_t^{\infty}\overline{G}(u)\mathrm{d}u+\int_0^t\int_{t-s}^{\infty}\overline{F}(u)\mathrm{d}u\mathrm{d}m(s), \qquad t\geqslant 0. \tag{14}$$

证 首先注意到 $N(t)+1$ 关于 $\{W_n, n\geqslant 1\}$ 是停时，与第一章定理 2.1 的证明相仿，我们有

$$\begin{aligned}\mathrm{E}[S_{N(t)+1}] &= \sum_{n=1}^{\infty} \mathrm{E}[W_n 1_{\{N(t)+1\geqslant n\}}] = \sum_{n=1}^{\infty} \mathrm{E}[W_n]\mathrm{P}(N(t)+1\geqslant n) \\ &= \mathrm{E}[W_1] + \mathrm{E}[W_2]\sum_{n=1}^{\infty}\mathrm{P}(N(t)\geqslant n) = \mu_G + \mu_F m(t) < \infty,\end{aligned}$$

由于 $A(t)+R(t)=\beta(t)\leqslant S_{N(t)+1}$，因此 $\mathrm{E}[R(t)]$ 为局部有界函数.

记 $B(t)=\mathrm{E}[R(t)]$. 先设 N 为更新过程，则因

$$\mathrm{E}[R(t)\mid W_1=s]=\begin{cases} s-t, & s>t, \\ \mathrm{E}[R(t-s)], & s\leqslant t,\end{cases}$$

所以

$$\begin{aligned}B(t) &= \int_t^{\infty}(s-t)\mathrm{d}F(s) + \int_0^t B(t-s)\mathrm{d}F(s) \\ &= \int_t^{\infty}\overline{F}(s)\mathrm{d}s + \int_0^t B(t-s)\mathrm{d}F(s).\end{aligned}$$

由定理 1.6 可解得

$$B(t)=\int_t^{\infty}\overline{F}(s)\mathrm{d}s + \int_0^{\infty}\int_{t-s}^{\infty}\overline{F}(u)\mathrm{d}u\mathrm{d}m(u),$$

即(14)成立.

对一般的延迟更新过程，与前面的证明相仿，

$$B(t)=\int_0^{\infty}\mathrm{E}[R(t)\mid W_1=s]\mathrm{d}G(s). \tag{15}$$

这时

$$\mathrm{E}[R(t)\mid W_1=s)=\begin{cases} s-t, & s>t, \\ g(t-s), & s\leqslant t,\end{cases}$$

其中的

$$g=h+h*m_F, \qquad h(t)=\int_t^{\infty}\overline{F}(s)\mathrm{d}s.$$

仍由 $g*G=h*(G+m_F*G)=h*m$ 可得(14)式. □

关于 $\mathrm{E}[\beta(t)]$ 的讨论留给读者进行(习题 2-3.).

定义 设 $\mu_F=\int_0^{\infty}[1-F(t)]\mathrm{d}t<\infty$，称

$$F_e(t)=\begin{cases}\dfrac{1}{\mu_F}\displaystyle\int_0^t[1-F(s)]\mathrm{d}s, & t\geqslant 0, \\ 0, & t<0\end{cases} \tag{16}$$

为 F 的平衡分布. 显然，平衡分布函数 F_e 总是连续的，不管原来的 F 是否连续.

若延迟更新过程 N 的首次更新间隔 W_1 的分布 $G=F_e$，则 N 又称为稳定更新过程.

在下一节中我们将会看到，稳定更新过程反映延迟更新过程经长时间发展变化后进入稳定态后的特性.

容易看出，指数分布的平衡分布就是自己，因此泊松过程是稳定更新过程.

定理 2.9 设延迟更新过程 N 满足条件：

$$\mu_F=\int_0^\infty \overline{F}(t)\mathrm{d}t<\infty,$$

则 N 为稳定更新过程的充要条件是

$$m(t)=\frac{t}{\mu_F},\qquad t\geqslant 0. \tag{17}$$

证 取拉普拉斯变换

$$\widetilde{m}(\lambda)=\int_0^\infty \mathrm{e}^{-\lambda t}\mathrm{d}m(t),\quad \widetilde{F}(\lambda)=\int_0^\infty \mathrm{e}^{-\lambda t}\mathrm{d}F(t),\quad \widetilde{G}(\lambda)=\int_0^\infty \mathrm{e}^{-\lambda t}\mathrm{d}G(t),$$

则由(1)式 $\widetilde{m}(\lambda)=\widetilde{G}(\lambda)+\widetilde{m}(\lambda)\widetilde{F}(\lambda)$，故

$$\widetilde{m}(\lambda)=\frac{\widetilde{G}(\lambda)}{1-\widetilde{F}(\lambda)}.$$

F_e 的拉普拉斯变换为

$$\widetilde{F}_e(\lambda)=\frac{1}{\mu_F}\int_0^\infty \mathrm{e}^{-\lambda t}(1-F(t))\mathrm{d}t=\frac{1}{\mu_F\lambda}-\frac{1}{\mu_F\lambda}\int_0^\infty \mathrm{e}^{-\lambda t}\mathrm{d}F(t)=\frac{1-\widetilde{F}(\lambda)}{\mu_F\lambda}.$$

由此，简单的计算表明，$G=F_e$ 等价于 $\widetilde{m}(\lambda)=1/\mu_F\lambda$，$m(t)=t/\mu_F$. □

定理 2.10 设 N 为稳定更新过程，则

1) 对一切 $t\geqslant 0$，$R(t)$的分布函数为 F_e；

2) N 有平稳增量，即对任意的 $n\geqslant 1$ 及 $t_1<t_2<\cdots<t_n$，$\{N(t+t_i)-N(t),1\leqslant i\leqslant n\}$ 的分布与 t 无关.

证 1) 由(11)式及 $m(t)=t/\mu_F$，对一切 $t\geqslant 0,x\geqslant 0$

$$\begin{aligned}\mathsf{P}(R(t)>x)&=\frac{1}{\mu_F}\int_{t+x}^\infty \overline{F}(u)\mathrm{d}u+\frac{1}{\mu_F}\int_0^t \overline{F}(t-s+x)\mathrm{d}s\\&=\frac{1}{\mu_F}\int_{t+x}^\infty \overline{F}(u)\mathrm{d}u+\frac{1}{\mu_F}\int_x^{t+x}\overline{F}(u)\mathrm{d}u=\frac{1}{\mu_F}\int_x^\infty \overline{F}(u)\mathrm{d}u.\end{aligned}$$

所以，$R(t)$的分布函数是 F_e.

2) 对任一固定的 $s>0$，令 $Y^{(s)}=N(t+s)-N(t)$，它是一个计数过程，我们要证明它还是延迟更新过程. 事实上，它的第一个等待时间是 $R(s)$，它的分布是 F_e，接下去的等待时间是 $W_{N(s)+2},W_{N(s)+3},\cdots$，需要证明的是 $R(s),W_{N(s)+2},W_{N(s)+3},\cdots$ 相互独立及 $W_{N(s)+2}$，$W_{N(s)+3},\cdots$为同分布 F 的.

对任意的 $n\geqslant 1$ 及实数 $x,x_1,\cdots,x_n$，

$$
\begin{aligned}
&\mathsf{P}(R(s)\leqslant x, W_{N(s)+2}\leqslant x_1,\cdots,W_{N(s)+n+1}\leqslant x_n)\\
=&\sum_{k=1}^{\infty}\mathsf{P}(R(s)\leqslant x, N(s)+1=k, W_{N(s)+2}\leqslant x_1,\cdots,W_{N(s)+n+1}\leqslant x_n)\\
=&\sum_{k=1}^{\infty}\mathsf{P}(S_k-s\leqslant x, S_{k-1}\leqslant s<S_k, W_{k+1}\leqslant x_1,\cdots,W_{k+n}\leqslant x_n)\\
=&\sum_{k=1}^{\infty}\mathsf{P}(S_k-s\leqslant x, S_{k-1}\leqslant s<S_k)F(x_1)\cdots F(x_n)\\
=&\sum_{k=1}^{\infty}\mathsf{P}(R(s)\leqslant x, N(s)+1=k)F(x_1)\cdots F(x_n)\\
=&\mathsf{P}(R(s)\leqslant x)F(x_1)\cdots F(x_n)=F_e(x)F(x_1)\cdots F(x_n).
\end{aligned}
$$

这样,我们证明了,两个延迟更新过程 $Y^{(s)}$ 与 N 的等待时间的分布完全相同,因此这两个过程同分布. □

习　题

2-1. 设 N 为更新过程,$t_0, t_1>0$,

$$
Y(t)=X(t+t_0)-X(t_0),\qquad Z(t)=Y(t+t_1)-Y(t_1),\qquad t\geqslant 0,
$$

则 $Y=\{Y(t),t\geqslant 0\}$ 与 $Z=\{Z(t),t\geqslant 0\}$ 均为延迟更新过程.

2-2. 试对更新过程 $\{N(t),t\geqslant 0\}$ 计算 $\mathsf{P}(R(t)>x\mid A(t)=s)$ 及 $\mathsf{P}(R(t)>x\mid A(t+x/2)=s)$.

对于泊松过程,$\mathsf{P}(R(t)>x\mid A(t+x)=s)=?$

2-3. 对延迟更新过程 N,证明:以概率 1

$$
\lim_{t\to\infty}\frac{A(t)}{t}=0.
$$

2-4. 设 N 为延迟更新过程.对一切 $t\geqslant 0$,以 H_t 记 $S_{N(t)}$ 的分布函数,则

$$
H_t(s)=\begin{cases}\overline{G}(t)+\int_0^s \overline{F}(t-y)\mathrm{d}m(y), & t>s\geqslant 0,\\ 1, & t\leqslant s.\end{cases}
$$

当 $t>s\geqslant 0$ 时,上式又可写为

$$
\begin{cases}H_t(0)=\overline{G}(t)+\overline{F}(t)m(0)=\overline{G}(t)+\overline{F}(t)\dfrac{G(0)}{1-F(0)},\\ \mathrm{d}H_t(s)=\overline{F}(t-s)\mathrm{d}m(s), & s>0.\end{cases}
$$

试利用对"t 前最后一个更新时刻"$S_{N(t)}$ 取条件概率的全概率公式

$$
\mathsf{P}(A(t)\geqslant x)=\int_0^t \mathsf{P}(A(t)\geqslant x\mid S_{N(t)}=s)\mathrm{d}H_t(s),
$$

$$
\mathsf{P}(R(t)>x)=\int_0^t \mathsf{P}(R(t)>x\mid S_{N(t)}=s)\mathrm{d}H_t(s),
$$

$$\mathsf{P}(\beta(t)>x)=\int_0^t \mathsf{P}(\beta(t)>x \mid S_{N(t)}=s)\mathrm{d}H_t(s),$$

导出 $\mathsf{P}(A(t)\geqslant x)$,$\mathsf{P}(R(t)>x)$及 $\mathsf{P}(\beta(t)>x)$的公式.

2-5. 设延迟更新过程 N 满足条件:$\mu_G=\mathsf{E}[W_1]<\infty$,$\mu_F=\mathsf{E}[W_2]<\infty$,则

$$\mathsf{E}[\beta(t)]=\int_t^\infty u\mathrm{d}G(u)+\int_0^t\int_{t-s}^\infty u\mathrm{d}F(u)\mathrm{d}m(s).$$

2-6. 若一个等待时间分布的平衡分布就是自身,则此分布必为指数分布. 由此可知,一个稳定更新过程又是更新过程当且仅当该过程是泊松过程.

2-7. 求平衡分布 F_e,若

1) F 为退化于 $c>0$ 点的退化分布;

2) F 为(0,1)上的均匀分布.

2-8. 对延迟更新过程 N,证明:$t\geqslant 0$,$x>0$ 时

$$\mathsf{P}(\bigcup_{n=1}^\infty\{S_n\in(t,t+x]\})=G(t+x)-G(t)+\int_0^t[F(t-s+x)-F(t-s)]\mathrm{d}m(s).$$

特别,对稳定更新过程,

$$\mathsf{P}(\bigcup_{n=1}^\infty\{S_n\in(t,t+x]\})=F_e(x).$$

2-9. 在Ⅰ型计数器(例 1.1)中,设信号来到间隔的分布为 F,信号使计数器关闭的时间的分布为 G,$\mu_F<\infty$,$\mu_G<\infty$,证明:

1) n_1 的分布为

$$\mathsf{P}(n_1=k)=\int_0^\infty [F^{*(k-1)}(t)-F^{*k}(t)]\mathrm{d}G(t),\qquad k\geqslant 1;$$

2) $\mathsf{E}[n_1]=1+\int_0^\infty m_F(t)\mathrm{d}G(t)$;

3) S_{n_1} 的分布为

$$\mathsf{P}(S_{n_1}\leqslant t)=\int_0^t G(s^-)[1-F(t-s)]\mathrm{d}m_F(s);$$

4) $\mathsf{E}[S_{n_1}]=\mu_F\ \mathsf{E}[n_1]$.

§4.3 更新定理

定理 3.1 (初等更新定理) 对延迟更新过程 $N=\{N(t),t\geqslant 0\}$成立

$$\lim_{t\to\infty}\frac{m(t)}{t}=\frac{1}{\mu_F}. \tag{1}$$

证 若 $\mu_G<\infty$,$\mu_F<\infty$,在定理 2.8 的证明中已知

$$t<\mathsf{E}[S_{N(t)+1}]=\mu_G+\mu_F m(t),$$

故

$$\underline{\lim_{t\to\infty}}\frac{m(t)}{t}\geqslant\frac{1}{\mu_F}. \tag{2}$$

另一方面,令

$$\widetilde{W}_n=\min(W_n,M),n\geqslant1,\mathsf{E}[\widetilde{W}_1]=v_M,\mathsf{E}[\widetilde{W}_n]=\mu_M,n\geqslant2,$$

$$\tilde{S}_n=\sum_{j=1}^n W_j,\qquad \widetilde{N}(t)=\sup_n\{n:\tilde{S}_n\leqslant t\},\quad \widetilde{m}(t)=\mathsf{E}[\widetilde{N}(t)],$$

则$\{\widetilde{N}(t),t\geqslant0\}$为延迟更新过程,以$\{\widetilde{W}_n,n\geqslant1\}$为等待时间序列.因 $\tilde{S}_{\widetilde{N}(t)}\leqslant t,\tilde{S}_{\widetilde{N}(t)+1}\leqslant t+M$,故

$$v_M+\mu_M\widetilde{m}(t)=\mathsf{E}[\tilde{S}_{\widetilde{N}(t)+1}]<t+M.$$

因此

$$\overline{\lim_{t\to\infty}}\frac{\widetilde{m}(t)}{t}\leqslant\frac{1}{\mu_M}.$$

又因

$$\widetilde{W}_n\leqslant W_n,\tilde{S}_n\leqslant S_n,\widetilde{N}(t)\geqslant N(t),\quad \widetilde{m}(t)\geqslant m(t),$$

所以

$$\overline{\lim_{t\to\infty}}\frac{m(t)}{t}\leqslant\frac{1}{\mu_M}.$$

令 $M\to\infty$,由于 $\mu_M\to\mu_F$,故得

$$\overline{\lim_{t\to\infty}}\frac{m(t)}{t}\leqslant\frac{1}{\mu_F}. \tag{3}$$

若 $\mu_F=\infty$,由(3)式即得(1)式.若 $\mu_F<\infty$,联合(2)式及(3)式,也得(1)式. □

$m(t)=\mathsf{E}[N(t)]$是$[0,t]$中 N 发生的更新次数的平均值,因此 $m(t)/t$ 是单位时间内 N 的平均更新次数.另一方面,μ_F 是从第二次更新起每一次更新的平均等待时间.因此,(1)式表明长时间之后单位时间内更新次数的平均值等于一次更新的平均等待时间.显然,长时间之后,第一次更新的平均时间对单位时间内更新次数的平均值不起作用.

引理 3.1 设 N 为延迟更新过程,则对一切 $t\geqslant0,a>0,k\geqslant1$,

$$\mathsf{P}([N(t+a)-N(t)]\geqslant k)\leqslant\mathsf{P}([\widetilde{N}(a)+1]\geqslant k), \tag{4}$$

其中 $\widetilde{N}$ 是等待时间分布为F 的更新过程.特别,对一切 $t\geqslant0$

$$m(t+a)-m(t)\leqslant m_F(a)+1. \tag{5}$$

证 令 $n_t=\inf\{n\geqslant1:S_n\in(t,t+a]\}$,则对 $k\geqslant1$

$$
\begin{aligned}
\mathsf{P}([N(t+a)-N(t)]\geqslant k) &= \mathsf{P}\Big(\Big[1+\sum_{n=n_t+1}^{\infty}1_{(t,t+a]}(S_n)\Big]\geqslant k, n_t<\infty\Big)\\
&= \mathsf{P}\Big(\Big[1+\sum_{n=n_t+1}^{\infty}1_{(t-S_{n_t},t+a-S_{n_t}]}(S_n-S_{n_t})\Big]\geqslant k, n_t<\infty\Big)\\
&\leqslant \mathsf{P}\Big(\Big[1+\sum_{n=n_t+1}^{\infty}1_{(0,a]}(S_n-S_{n_t})\Big]\geqslant k, n_t<\infty\Big)\\
&\leqslant \mathsf{P}\Big(\Big[1+\sum_{n=2}^{\infty}1_{(0,a]}(S_n-S_1)\Big]\geqslant k\Big)=\mathsf{P}([1+\tilde{N}(a)]\geqslant k),
\end{aligned}
$$

因为 n_t 关于 $\{S_n, n\geqslant 1\}$ 是停时，在 $n_t<\infty$ 的条件下，$\{S_{n+n_t}-S_{n_t}, n\geqslant 1\}$ 的条件分布与 $\{S_n, n\geqslant 1\}$ 的分布相同(参见第一章定理 2.1)，从而

$$
\begin{aligned}
m(t+a)-m(t) &= \mathsf{E}[N(t+a)-N(t)] = \sum_{k=1}^{\infty}\mathsf{P}([N(t+a)-N(t)]\geqslant k)\\
&\leqslant \sum_{k=1}^{\infty}\mathsf{P}([1+\tilde{N}(a)]\geqslant k) = \mathsf{E}[1+\tilde{N}(a)] = 1+m_F(a). \qquad \square
\end{aligned}
$$

引理 3.2 设在延迟更新过程中，$G(t)=F^{(c)}(t)$，其中

$$
F^{(c)}(t)=\begin{cases}0, & t<0,\\ \dfrac{\int_0^t \bar{F}(s)\mathrm{d}s}{\int_0^c \bar{F}(s)\mathrm{d}s}, & 0\leqslant t<c, \quad c>0,\\ 1, & t\geqslant c,\end{cases} \tag{6}
$$

则对一切 $a>0$

$$
m(t+a)-m(t)\leqslant \frac{a}{\int_0^c \bar{F}(s)\mathrm{d}s}. \tag{7}
$$

证 记 $\gamma(t)=\int_0^t \bar{F}(s)\mathrm{d}s$. 易见，由于 $F(0)<1, \gamma(t)>0, t>0$. 令

$$
\phi(t)=\frac{\gamma(t)}{\gamma(c)}-G(t)=\begin{cases}0, & t<c,\\ \dfrac{1}{\gamma(c)}\int_c^t \bar{F}(s)\mathrm{d}s, & t\geqslant c,\end{cases}
$$

则 $\phi(t)$ 为非负连续单调增加函数. 另一方面，

$$
\begin{aligned}
t &= \int_0^t \bar{F}(s)\mathrm{d}s+\int_0^t F(s)\mathrm{d}s\\
&= \gamma(t)+\int_0^t F(t-s)\mathrm{d}s=\gamma(t)+\int_0^t (t-s)\mathrm{d}F(s), \quad t\geqslant 0.
\end{aligned}
$$

因此，更新方程

$$B(t)=\gamma(t)+\int_0^t B(t-s)\mathrm{d}F(s) \tag{8}$$

的局部有界解为 $B(t)=t$(实际上,由定理 2.9 也可得出这结论). 但是(8)的解为 $B(t)=\gamma(t)+\int_0^t \gamma(t-s)\mathrm{d}m_F(s)$,因此

$$\gamma(t)+\int_0^t \gamma(t-s)\mathrm{d}m_F(s)=t, \qquad t\geqslant 0.$$

现在,

$$\begin{aligned} m(t) &= G(t)+\int_0^t G(t-s)\mathrm{d}m_F(s) \\ &= \frac{1}{\gamma(c)}\left[\gamma(t)+\int_0^t \gamma(t-s)\mathrm{d}m_F(s)\right]-\left[\phi(t)+\int_0^t \phi(t-s)\mathrm{d}m_F(s)\right] \\ &= \frac{1}{\gamma(c)}-[\phi+\phi*m_F](t). \end{aligned}$$

由于 $\phi+\phi*m_F$ 也单调增加,所以

$$m(t+a)-m(t)=\frac{a}{\gamma(c)}-\{[\phi+\phi*m_F](t+a)-[\phi+\phi*m_F](t)\}\leqslant\frac{a}{\gamma(c)},$$

即(7)式成立. □

在进一步介绍更新函数的渐近性质时,必须区分随机变量的分布的两种不同类型.

定义 设 H 为随机变量 X 的分布函数,

$$\mathrm{ssup}(H)=\{t:\forall\,\varepsilon>0, H(t+\varepsilon)-H(t-\varepsilon)>0\}$$

称为分布 H 的**支集**, $\mathrm{ssup}(H)$ 中的点称为 H 的**增长点**. 若存在 $\mathrm{d}>0$ 使

$$\mathrm{ssup}(H)\subset\{n\mathrm{d}: n \text{ 为整数}\}, \tag{9}$$

即

$$\sum_{n=-\infty}^{\infty}\mathrm{P}(X=n\mathrm{d})=1,$$

则称分布 H 为**格子点分布**, X 为**格子点变量**,满足(9)式的最大正数 d 称为格子点分布 H 的步长. 若无满足(9)的 d,则分布 H 称为**非格子点分布**. 显然,格子点分布是离散型分布,而连续型分布是非格子点分布.

引理 3.3 设 N 及 N' 为两个相互独立的更新过程,有共同的等待时间分布 F,且 F 非格子点分布,则对一切实数 x 成立

$$\forall\,\varepsilon>0, \exists\, i,j\geqslant 1 \text{ 使得 } \mathrm{P}(|x+S_i-S'_j|\leqslant\varepsilon)>0, \tag{10}$$

其中 $\{S_n, n\geqslant 1\}$ 与 $\{S'_n, n\geqslant 1\}$ 分别是 N 与 N' 的更新时刻.

证 以 D 记具有性质(10)的实数 x 全体,则 D 有下列性质:

1° $a\in D\Rightarrow -a\in D$,即 D 是对称集.

这是因为 $\mathrm{P}(|-a+S_j-S_i'|\leqslant\varepsilon)=\mathrm{P}(|a+S_i-S_j'|\leqslant\varepsilon)$.

2° $a,b\in D\Rightarrow a+b\in D$.

这是因为对任意的 $\varepsilon>0$,若有 $i,j,m,n\geqslant1$ 使

$$\mathrm{P}(|a+S_i-S'_j|\leqslant\varepsilon/2)>0,\quad \mathrm{P}(|b+S_m-S'_n|\leqslant\varepsilon/2)>0,$$

则

$$\begin{aligned}&\mathrm{P}(|a+b+S_{i+m}-S'_{j+n}|\leqslant\varepsilon)\\ &\geqslant\mathrm{P}(|a+S_i-S'_j|\leqslant\varepsilon/2)\mathrm{P}(|b+(S_{i+m}-S_i)-(S'_{j+n}-S'_j)|\leqslant\varepsilon/2)\\ &=\mathrm{P}(|a+S_i-S'_j|\leqslant\varepsilon/2)\mathrm{P}(|b+S_m-S'_n|\leqslant\varepsilon/2)>0.\end{aligned}$$

3° $a_n\in D,a_n\to a\Rightarrow a\in D$,即 D 是闭集.

这是因为对任意的 $\varepsilon>0$,有 a_n 使 $|a-a_n|\leqslant\varepsilon/2$,从而有 $i,j\geqslant1$ 使

$$\mathrm{P}(|a_n+S_i-S'_j|\leqslant\varepsilon/2)>0,$$

$$\mathrm{P}(|a+S_i-S'_j|\leqslant\varepsilon)\geqslant\mathrm{P}(|a_n+S_i-S'_j|\leqslant\varepsilon/2)>0.$$

4° $(0,\infty)\cap D$ 不是空集.

这是因为若 $x\notin D$,$\exists\,\varepsilon_x>0$ 使对一切 $i,j\geqslant1$,$\mathrm{P}(|x+S_i-S'_j|\leqslant\varepsilon_x)=0$,即

$$\mathrm{P}\Big(\bigcup_{i,j=1}^{\infty}\{S'_j-S_i\in[x-\varepsilon_x,x+\varepsilon_x]\}\Big)=0.$$

若$(0,\infty)\cap D$ 是空集,则对一切 $0<a<b$,由有限覆盖定理可知

$$\mathrm{P}\Big(\bigcup_{i,j=1}^{\infty}\{S'_j-S_i\in[a,b]\}\Big)=0.$$

进而可得

$$\mathrm{P}\Big(\bigcup_{i,j=1}^{\infty}\{S'_j-S_i\in(0,\infty)\}\Big)=0,\quad \mathrm{P}\Big(\bigcup_{i,j=1}^{\infty}\{S'_j-S_i\in(-\infty,0)\}\Big)=0,$$

$$\mathrm{P}\Big(\bigcup_{i,j=1}^{\infty}\{S_i=S'_j\}\Big)=1.$$

这显然是不可能的.

现在记 $c=\inf\{x>0:x\in D\}$.若 $c>0$,则 $D=\{0,\pm c,\pm2c,\cdots\}$.这是因为每个 $a\in D$ 可写成$a=nc+b$,$0\leqslant b<c$,其中 n 为整数.若 $b>0$,则 $b\in D$,与 c 的定义矛盾.由于 F 是非格子点分布,必有 $x\notin D$,$x\in\mathrm{ssup}(F)$,从而对任意的 $\varepsilon>0$

$$\begin{aligned}&\mathrm{P}(|x+S_2-S'_1|\leqslant\varepsilon)\\ &\geqslant\mathrm{P}(|x+W_1|\leqslant\varepsilon/3)\mathrm{P}(|x+W_2|\leqslant\varepsilon/3)\mathrm{P}(|x+W'_1|\leqslant\varepsilon/3)>0,\end{aligned}$$

$x\in D$,得出矛盾.所以 $c=0$.因此有一列 $c_n\downarrow0$,$c_n\in D$,

$$D\supset\bigcup_n\{0,\pm c_n,\pm2c_n,\cdots\}.$$

所以 D 在$(-\infty,\infty)$中稠密,但 D 是闭集,故 $D=(-\infty,\infty)$. □

定理 3.2(布莱克威尔(Blackwell)定理) 若延迟更新过程 N 的等待时间分布 F 为非格子点分布,则对一切 $a>0$

$$\lim_{t\to\infty}[m(t+a)-m(t)]=\frac{a}{\mu_F}. \tag{11}$$

证 若 N 为稳定更新过程,(11)是显然的.对一般的延迟更新过程,我们要找一个稳定更新过程“逼近”它.先用最简单的独立耦合,设 N' 是与 N 相互独立的延迟更新过程,其等待时间序列为 $\{W'_n, n\geqslant1\}$,更新时刻序列为 $\{S'_n, n\geqslant1\}$, W'_1 的分布为 G', W'_2 的分布为 F.

对任意的 $\varepsilon>0$,令

$$\begin{aligned}
\upsilon(\varepsilon)&=\inf\{n\geqslant1:\exists\, m\geqslant1 \text{ 使得 } |S_n-S'_m|\leqslant\varepsilon\},\\
\upsilon'(\varepsilon)&=\inf\{m\geqslant1:\exists\, n\geqslant1 \text{ 使得 } |S'_m-S_n|\leqslant\varepsilon\},\\
\tau(\varepsilon)&=\sup\{S_n:n\geqslant1,\exists\, m\geqslant1 \text{ 使得 } |S_n-S'_m|\leqslant\varepsilon\},
\end{aligned}$$

我们要证明:或者

$$\mathsf{P}(\tau(\varepsilon)<\infty)=1,\qquad \forall\,\varepsilon>0,G,G', \tag{12}$$

或者

$$\mathsf{P}(\upsilon(\varepsilon)<\infty)=1,\qquad \forall\,\varepsilon>0,G,G'. \tag{13}$$

设(13)式不成立,我们要证(12)式成立.这时存在 $\varepsilon>0,G,G'$ 使得 $\mathsf{P}(\upsilon(\varepsilon)=\infty)>0$,即

$$\begin{aligned}
0&<\mathsf{P}(|S_n-S'_m|>\varepsilon,\forall\, n,m\geqslant1)\\
&=\int\mathsf{P}(|S_n-S'_m|>\varepsilon,\forall\, n,m\geqslant1\mid S_1-S'_1=y)\mathrm{d}H(y),
\end{aligned}$$

其中 H 是 $S_1-S'_1$ 的分布.因此存在 $y\in(-\infty,\infty)$ 使得

$$\mathsf{P}(|y+\bar{S}_n-\bar{S}'_m|>\varepsilon,\forall\, n,m\geqslant1)>0, \tag{14}$$

这里 $\{\bar{S}_n=S_{n+1}-S_1,n\geqslant1\}$ 与 $\{\bar{S}'_m=S'_{m+1}-S'_1,m\geqslant1\}$ 是两个独立的有共同等待时间分布 F 的更新过程的更新时刻序列.取 $\delta=\varepsilon/3$,由引理 3.3,存在 $i,j\geqslant1$ 使得

$$\mathsf{P}(A)>0,\quad A=\{|\bar{S}_i-\bar{S}'_j-y|\leqslant\delta\}. \tag{15}$$

由(14)得

$$\mathsf{P}(B)>0,\quad B=\{|y+(\bar{S}_n-\bar{S}_i)-(\bar{S}'_m-\bar{S}'_j)|>3\delta,\forall\, n>i,m>j\}. \tag{16}$$

A 与 B 相互独立,且

$$AB\subset C=\{|\bar{S}_n-\bar{S}'_m|>2\delta,\forall\, n>i,m>j\}. \tag{17}$$

由(15)与(16)得

$$p=\mathsf{P}(C)\geqslant\mathsf{P}(AB)=\mathsf{P}(A)\mathsf{P}(B)>0. \tag{18}$$

注意(14)—(18)与第一个等待时间 W_1,W'_1 无关.现在我们仍设 G 与 G' 是任意的.令

$$\upsilon_1=\upsilon(\delta),\qquad \upsilon'_1=\upsilon'(\delta),$$

$$v_{k+1}=\inf\{n>v_k:\exists\, m>v'_k\text{使得}|S_n-S'_m|\leqslant\delta\},$$
$$v'_{k+1}=\inf\{m>v'_k:\exists\, n>v_k\text{ 使得}|S_n-S'_m|\leqslant\delta\},\quad k\geqslant 1.$$

若 $v_k<\infty$(从而 $v'_k<\infty$),则$\{S_{v_k+n}-S_{v_k},n\geqslant 1\}$与$\{S'_{v'_k+n}-S'_{v'_k},n\geqslant 1\}$是两个独立的有共同等待时间分布 F 的更新过程的更新时刻序列.令

$$C_k=\{v_k<\infty\}\cap\{|(S_{v_k+n}-S_{v_k})-(S'_{v'_k+m}-S'_{v'_k})|>2\delta,\forall\, n\geqslant i,m>j\},$$

则(参见第一章定理2.1)

$$\mathrm{P}(C_k|v_k<\infty)=\mathrm{P}(C)=p>0.$$

易见

$$\{v_{k+i+1}<\infty\}\subset\{v_k<\infty\}\cap\bar{C}_k,$$
$$\mathrm{P}(v_{k+i+1}<\infty)\leqslant\mathrm{P}(v_k<\infty)(1-p),$$
$$\mathrm{P}(v_{n(i+1)+1}<\infty)\leqslant(1-p)^n,\qquad n\geqslant 1.$$

然而$\{v_{n(i+1)+1}<\infty\}\downarrow\{\tau(\delta)=\infty\}$,故得 $\mathrm{P}(\tau(\delta)<\infty)=1$.

对任意的 $n,m\geqslant 1$,

$$\{|S_n-S'_m|\leqslant 2\delta\}=\{|\delta+S_n-S'_m|\leqslant\delta\}\cup\{|S_n-S'_m-\delta|\leqslant\delta\},$$

因此 $\tau(2\delta)=\max\{\tau_\delta(\delta)-\delta,\tau'_\delta(\delta)+\delta\}$,其中 $\tau_\delta(\delta)$是将 W_1 换成 $W_1+\delta$ 后与$\tau(\delta)$相应的随机变量,$\tau'_\delta(\delta)$是将 W'_1 换成 $W'_1+\delta$ 后相应的随机变量.由已证得的结果,$\mathrm{P}(\tau_\delta(\delta)=\infty)=\mathrm{P}(\tau'_\delta(\delta)=\infty)=0$,因此 $\mathrm{P}(\tau(2\delta)=\infty)=0$,进而对一切 $k\geqslant 1$,$\mathrm{P}(\tau(2^k\delta)=\infty)=0$.由于 $\tau(\epsilon)$随 ϵ 的增加而增加,所以对一切 $\epsilon>0$,$\mathrm{P}(\tau(\epsilon)=\infty)=0$,即(12)式成立.

现在,我们分别就(12)或(13)成立的两种情况证明(11)式.

先设(12)式成立.这时取 $G'=G$.记 $A_t=\{N(t+a)-N(t)\geqslant 1\}$,则

$$[\mathrm{P}(A_t)]^2=\mathrm{P}(N(t+a)-N(t)\geqslant 1,N'(t+a)-N'(t)\geqslant 1)\leqslant\mathrm{P}(\tau(a)>t),\tag{19}$$

因为 $N(t+a)-N(t)\geqslant 1$ 时有 $S_i\in(t,t+a]$,$N'(t+a)-N'(t)\geqslant 1$ 时有 $S'_j\in(t,t+a]$,从而$|S_i-S'_j|\leqslant a$,$\tau(a)\geqslant S_i>t$.由(12)及(19)得 $\lim_{t\to\infty}\mathrm{P}(A_t)=0$.由引理3.1,对任意的 $k\geqslant 1$

$$\begin{aligned}0\leqslant m(t+a)-m(t)&=\mathrm{E}[(N(t+a)-N(t))1_{A_t}]\\&\leqslant k\mathrm{P}(A_t)+\mathrm{E}[(N(t+a)-N(t))1_{\{[N(t+a)-N(t)]\geqslant k\}}]\\&\leqslant k\mathrm{P}(A_t)+\mathrm{E}[(\widetilde{N}(a)+1)1_{\{[\widetilde{N}(a)+1]\geqslant k\}}].\end{aligned}$$

在上式中先令 $t\to\infty$,再令 $k\to\infty$得

$$\lim_{t\to\infty}[m(t+a)-m(t)]=0.\tag{20}$$

这时若 $\mu_F<\infty$,取 $G=F_e$,则 $m(t+a)-m(t)=a/\mu_F>0$,与(20)式矛盾.因此 $\mu_F=\infty$,故(11)式成立.

再设(13)式成立.这时对一切 $\epsilon>0$,$\mathrm{P}(v(\epsilon)<\infty)=\mathrm{P}(v'(\epsilon)<\infty)=1$.下面用一种非独立

耦合,定义

$$W''_n=\begin{cases}W'_n, & n\leqslant \upsilon'(\varepsilon),\\ W_{\upsilon(\varepsilon)+n-\upsilon'(\varepsilon)}, & n>\upsilon'(\varepsilon),\end{cases}$$

则$\{W''_n,n\geqslant 1\}$与$\{W'_n,n\geqslant 1\}$同分布.事实上,对任意的 $n\geqslant 1$ 与实数 $x_1,\cdots,x_n$,

$$\begin{aligned}\mathsf{P}(W''_1\leqslant x_1,\cdots,W''_n\leqslant x_n)&=\sum_{i,j=1}^{\infty}\mathsf{P}(\upsilon(\varepsilon)=i,\upsilon'(\varepsilon)=j,W''_1\leqslant x_1,\cdots,W''_n\leqslant x_n)\\&=\sum_{j=n}^{\infty}\sum_{i=1}^{\infty}\mathsf{P}(\upsilon(\varepsilon)=i,\upsilon'(\varepsilon)=j,W'_1\leqslant x_1,\cdots,W'_n\leqslant x_n)+\sum_{j=1}^{n-1}\sum_{i=1}^{\infty}\mathsf{P}(\upsilon(\varepsilon)=i,\\&\quad\upsilon'(\varepsilon)=j,W'_1\leqslant x_1,\cdots,W'_j\leqslant x_j,W_{i+1}\leqslant x_{j+1},\cdots,W_{i+n-j}\leqslant x_n).\end{aligned}\tag{21}$$

事件$\{\upsilon(\varepsilon)=i,\upsilon'(\varepsilon)=j\}=\{|S_i-S'_j|\leqslant\varepsilon,|S_n-S'_m|>\varepsilon,n<i,m<j\}$发生与否由$\{W_1,\cdots,W_i,W'_1,\cdots,W'_j\}$完全决定.当 $n\geqslant j+1$ 时

$$\begin{aligned}&\mathsf{P}(\upsilon(\varepsilon)=i,\upsilon'(\varepsilon)=j,W'_1\leqslant x_1,\cdots,W'_j\leqslant x_j,W_{i+1}\leqslant x_{j+1},\cdots,W_{i+n-j}\leqslant x_n)\\&\quad=\mathsf{P}(\upsilon(\varepsilon)=i,\upsilon'(\varepsilon)=j,W'_1\leqslant x_1,\cdots,W'_j\leqslant x_j)\mathsf{P}(W_{i+1}\leqslant x_{j+1},\cdots,W_{i+n-j}\leqslant x_n)\\&\quad=\mathsf{P}(\upsilon(\varepsilon)=i,\upsilon'(\varepsilon)=j,W'_1\leqslant x_1,\cdots,W'_j\leqslant x_j)\mathsf{P}(W'_{j+1}\leqslant x_{j+1},\cdots,W'_n\leqslant x_n)\\&\quad=\mathsf{P}(\upsilon(\varepsilon)=i,\upsilon'(\varepsilon)=j,W'_1\leqslant x_1,\cdots,W'_n\leqslant x_n).\end{aligned}\tag{22}$$

将(22)代入(21)即得

$$\mathsf{P}(W''_1\leqslant x_1,\cdots,W''_n\leqslant x_n)=\mathsf{P}(W'_1\leqslant x_1,\cdots,W'_n\leqslant x_n).$$

设延迟更新过程 N'' 以$\{W''_n,n\geqslant 1\}$为等待时间序列,则 N'' 与 N' 同分布.因为$|S_{\upsilon(\varepsilon)}-S'_{\upsilon'(\varepsilon)}|\leqslant\varepsilon$,而 N''在$S'_{\upsilon'(\varepsilon)}$之后与 N 与$S_{\upsilon(\varepsilon)}$之后的等待时间相同,所以若 $S_{\upsilon(\varepsilon)}\leqslant t$,则

$$N''(t+a-\varepsilon)-N''(t+\varepsilon)\leqslant N(t+a)-N(t)\leqslant N''(t+a+\varepsilon)-N''(t-\varepsilon).$$

记 $B_t=\{S_{\upsilon(\varepsilon)}>t\}$,我们有

$$N(t+a)-N(t)\geqslant[N''(t+a-\varepsilon)-N''(t+\varepsilon)]1_{B_t},$$

$$m(t+a)-m(t)\geqslant m''(t+a-\varepsilon)-m''(t+\varepsilon)-\mathsf{E}[(N''(t+a)-N''(t))1_{B_t}].\tag{23}$$

$$N(t+a)-N(t)\leqslant N''(t+a+\varepsilon)-N''(t-\varepsilon)+[N(t+a)-N(t)]1_{B_t},$$

$$m(t+a)-m(t)\leqslant m''(t+a+\varepsilon)-m''(t-\varepsilon)+\mathsf{E}[(N(t+a)-N(t))1_{B_t}].\tag{24}$$

由引理 3.1,对任意的 $k\geqslant 1$,

$$\begin{aligned}\mathsf{E}[(N(t+a)-N(t))1_{B_t}]&\leqslant k\mathsf{P}(B_t)+\mathsf{E}[(N(t+a)-N(t))1_{\{[N(t+a)-N(t)]\geqslant k\}}]\\&\leqslant k\mathsf{P}(B_t)+\mathsf{E}[(\widetilde{N}(a)+1)1_{\{[\widetilde{N}(a)+1]\geqslant k\}}].\end{aligned}$$

由于$\lim_{t\to\infty}\mathsf{P}(B_t)=0$,在上式中先令 $t\to\infty$,再令 $k\to\infty$即得

$$\lim_{t\to\infty}\mathsf{E}[(N(t+a)-N(t))1_{B_t}]=0.\tag{25}$$

同理可证

$$\lim_{t\to\infty} \mathrm{E}[(N''(t+a)-N''(t))1_{B_t}]=0. \tag{26}$$

若 $\mu_F<\infty$,取 $G'=F_e$,则 $m''(t)=t/\mu_F$.由(23)—(26),

$$\overline{\lim_{t\to\infty}}\left|m(t+a)-m(t)-\frac{a}{\mu_F}\right|\leqslant\frac{2\varepsilon}{\mu_F}.$$

在上式中令 $\varepsilon\to 0$ 即得(11).

若 $\mu_F=\infty$,取 $G'=F^{(c)}$,$c>0$(见引理 3.2).由(24),(25)及引理 3.2,

$$\overline{\lim_{t\to\infty}}[m(t+a)-m(t)]\leqslant\frac{a+2\varepsilon}{\gamma(c)}.$$

在上式中令 $\varepsilon\to 0$,$c\to\infty$ 即得(11). □

初等更新定理是布莱克威尔定理的推论(习题 2-1.).事实上,布莱克威尔定理比初等更新定理精细得多.为了进一步扩展布莱克威尔定理的结论,我们引入下面的可积性概念.

定义 设 h 为$[0,\infty)$上的非负函数.对任一 $a>0$,令

$$\underline{m}_n(a)=\inf\{h(t):(n-1)a\leqslant t\leqslant na\},\quad \underline{m}(a)=\sum_{n=1}^{\infty}a\,\underline{m}_n(a);$$

$$\overline{m}_n(a)=\sup\{h(t):(n-1)a\leqslant t\leqslant na\},\quad \overline{m}(a)=\sum_{n=1}^{\infty}a\,\overline{m}_n(a).$$

若对一切 $a>0$,$\overline{m}(a)<\infty$

$$\lim_{a\to 0}\underline{m}(a)=\lim_{a\to 0}\overline{m}(a), \tag{27}$$

且这极限值有限,则称 h 在$[0,\infty)$上直接黎曼(Riemann)可积(直接 R 可积).

引理 3.4 设 $h\geqslant 0$ 在$[0,\infty)$上直接 R 可积,则 h 在$[0,\infty)$上黎曼可积,且(27)式中的极限即 $\int_0^\infty h(t)\mathrm{d}t$.

证 事实上,对任意的 $T>0$,取 $a_k=T/k$,则

$$0=\lim_k[\overline{m}(a_k)-\underline{m}(a_k)]\geqslant\lim_k a_k\sum_{n=1}^{k}[\overline{m}_n(a_k)-\underline{m}_n(a_k)]\geqslant 0,$$

$$\lim_k\{a_k\sum_{n=1}^{k}[\overline{m}_n(a_k)-\underline{m}_n(a_k)]\}=0.$$

这即表明 h 在$[0,T]$上黎曼可积,且

$$\lim_k\{a_k\sum_{n=1}^{k}\overline{m}_n(a_k)\}=\lim_k\{a_k\sum_{n=1}^{k}\underline{m}_n(a_k)\}=\int_0^T h(t)\mathrm{d}t.$$

现在取 $a_k=1/2^k$,则对固定的正整数 M,

$$\lim_{k\to\infty}\{a_k\sum_{n=1}^{2^kM}\overline{m}_n(a_k)\}=\int_0^M h(t)\mathrm{d}t,$$

$$\overline{m}(a_k)-a_k\sum_{n=1}^{2^kM}\overline{m}_n(a_k)=a_k\sum_{n=2^kM+1}^{\infty}\overline{m}_n(a_k)\leqslant\sum_{n=M+1}^{\infty}\overline{m}_n(1)\to 0,\quad M\to\infty.$$

$$|\overline{m}(a_k)-\int_0^M h(t)\mathrm{d}t|\leqslant\sum_{n=M+1}^{\infty}\overline{m}_n(1)+|a_k\sum_{n=1}^{2^kM}\overline{m}_n(a_k)-\int_0^M h(t)\mathrm{d}t|.$$

在上式中先令 $k\to\infty$,再令 $M\to\infty$,即得

$$\lim_{a\to 0}\overline{m}(a)=\lim_{k\to\infty}\overline{m}(a_k)=\int_0^\infty h(t)\mathrm{d}t. \qquad \square$$

不难举出在$[0,\infty)$上黎曼可积,但在$[0,\infty)$上不直接R可积的非负函数的例子.例如在$[0,\infty)$上黎曼可积的函数在$[0,\infty)$上甚至可以是无界的,但对直接R可积的函数 h, $\lim_{t\to\infty}h(t)=0$.下面的引理给出最有用的直接R可积的判别法.

引理 3.5 设 h 为$[0,\infty)$上的非负单调下降函数,且

$$\int_0^\infty h(t)\mathrm{d}t<\infty,$$

则 h 在$[0,\infty)$上直接R可积.

证 对任一 $a>0$,这时

$$\underline{m}_n(a)=h(na),\qquad \overline{m}_n(a)=h((n-1)a),$$

$$a\sum_{n=1}^{\infty}\underline{m}_n(a)=a\sum_{n=1}^{\infty}\underline{h}(na)\geqslant\int_a^\infty h(t)\mathrm{d}t,$$

$$a\sum_{n=1}^{\infty}\overline{m}_n(a)=a\sum_{n=1}^{\infty}h((n-1)a)\leqslant ah(0)+\int_0^\infty h(t)\mathrm{d}t<\infty,$$

从而有

$$\lim_{a\to 0}\{a\sum_{n=1}^{\infty}\underline{m}_n(a)\}=\lim_{a\to 0}\{a\sum_{n=1}^{\infty}\overline{m}_n(a)\}=\int_0^\infty h(t)\mathrm{d}t. \qquad \square$$

引进直接R可积性条件是为了保证下列关键更新定理成立,它是更新过程理论中最重要的极限定理.

定理 3.3(关键更新定理) 设延迟更新过程 N 的等待时间分布 F 为非格子点分布,h 为$[0,\infty)$上直接R可积的非负函数,则

$$\lim_{t\to\infty}\int_0^t h(t-s)\mathrm{d}m(s)=\frac{1}{\mu_F}\int_0^\infty h(s)\mathrm{d}s. \tag{28}$$

证 对任意的 $a>0$,令

$$h'_a(t)=\underline{m}_n(a),\qquad (n-1)a\leqslant t<na,\quad n\geqslant 1,$$

$$h''_a(t)=\overline{m}_n(a),\qquad (n-1)a\leqslant t<na,\quad n\geqslant 1.$$

易见,对一切 $k\geqslant 1, t>ka$,

$$\sum_{n=1}^{k}\underline{m}_n(a)[m(t-(n-1)a)-m(t-na)]\leqslant\int_0^t h'_a(t-s)\mathrm{d}m(s)$$

$$\leqslant \sum_{n=1}^{k} \underline{m}_n(a)[m(t-(n-1)a)-m(t-na)]+C(a)\sum_{n=k+1}^{\infty}\underline{m}_n(a), \tag{29}$$

其中 $C(a)=\sup_{t\geqslant 0}[m(t)-m(t-a)]$(由引理 3.1$C(a)<\infty$). 在(29)中先令 $t\to\infty$ 并利用(11)式,然后令 $k\to\infty$ 可得

$$\lim_{t\to\infty}\int_0^t h'_a(t-s)\mathrm{d}m(s)=\frac{1}{\mu_F}\underline{m}(a). \tag{30}$$

同理可证

$$\lim_{t\to\infty}\int_0^t h''_a(t-s)\mathrm{d}m(s)=\frac{1}{\mu_F}\overline{m}(a). \tag{31}$$

易见

$$\int_0^t h'_a(t-s)\mathrm{d}m(s)\leqslant\int_0^t h(t-s)\mathrm{d}m(s)\leqslant\int_0^t h''_a(t-s)\mathrm{d}m(s).$$

令 $t\to\infty$,由(30)及(31)得

$$\frac{1}{\mu_F}\underline{m}(a)\leqslant\varliminf_{t\to\infty}\int_0^t h(t-s)\mathrm{d}m(s)\leqslant\varlimsup_{t\to\infty}\int_0^t h(t-s)\mathrm{d}m(s)\leqslant\frac{1}{\mu_F}\overline{m}(a).$$

再令 $a\to 0$ 即得(28)式. □

推论 设 $b(t)\geqslant 0$ 在$[0,\infty)$上直接 R 可积,$B(t)$为更新方程

$$B(t)=b(t)+\int_0^t B(t-s)\mathrm{d}F(s)$$

的解,则

$$\lim_{t\to\infty}B(t)=\frac{1}{\mu_F}\int_0^\infty b(t)\mathrm{d}t.$$

证 由关键更新定理直接推得,因为

$$B(t)=b(t)+\int_0^t b(t-s)\mathrm{d}m_F(s),$$

$$\lim_{t\to\infty}b(t)=0. \quad \square$$

定理 3.4 设延迟更新过程 N 的等待时间分布 F 为非格子点分布,且 $\mu_F<\infty$,则

$$\lim_{t\to\infty}\mathrm{P}(A(t)\geqslant x)=\frac{1}{\mu_F}\int_x^\infty \overline{F}(t)\mathrm{d}t, \qquad x\geqslant 0, \tag{32}$$

$$\lim_{t\to\infty}\mathrm{P}(R(t)>x)=\frac{1}{\mu_F}\int_x^\infty \overline{F}(t)\mathrm{d}t, \qquad x\geqslant 0, \tag{33}$$

$$\lim_{t\to\infty}\mathrm{P}(\beta(t)>x)=\frac{1}{\mu_F}\int_x^\infty t\mathrm{d}F(t), \qquad x\geqslant 0. \tag{34}$$

证 我们有

$$\mathrm{P}(A(t)\geqslant x)=\overline{G}(t)1_{\{t\geqslant x\}}+\int_0^t \overline{F}(t-s)1_{\{t-s\geqslant x\}}\mathrm{d}m(s),$$

$\overline{F}(t)1_{\{t\geqslant x\}}$为非负单调下降函数,且

$$\int_0^\infty \overline{F}(t)1_{\{t\geqslant x\}}\mathrm{d}t=\int_x^\infty \overline{F}(t)\mathrm{d}t<\infty.$$

由引理 3.5,$\overline{F}(t)1_{\{t\geqslant x\}}$直接 R 可积.由关键更新定理即得(32)式.

我们有

$$\mathrm{P}(R(t)>x)=\overline{G}(t+x)+\int_0^t \overline{F}(t+x-s)\mathrm{d}m(s),$$

$\overline{F}(t+x)$为非负单调下降函数,且

$$\int_0^\infty \overline{F}(t+x)\mathrm{d}t=\int_x^\infty \overline{F}(t)\mathrm{d}t<\infty.$$

由引理 3.5,$\overline{F}(t+x)$直接 R 可积.由关键更新定理即得(33)式.(33)式也可利用 $R(t)$与 $A(t)$的关系式(第二节(6)式)及(32)式得到.

我们有

$$\mathrm{P}(\beta(t)>x)=\overline{G}(\max(x,t))+\int_0^t \overline{F}(\max(x,(t-s))\mathrm{d}m(s),$$

$\overline{F}(\max(x,t))$是非负单调下降函数,且

$$\int_0^\infty \overline{F}(\max(x,t))\mathrm{d}t=\int_0^x \overline{F}(t)\mathrm{d}t+\int_x^\infty \overline{F}(t)\mathrm{d}t$$
$$=x[1-F(x)]+\int_x^\infty \overline{F}(t)\mathrm{d}t=\int_x^\infty t\mathrm{d}F(t)<\infty.$$

由引理 3.5,$\overline{F}(\max(x,t))$直接 R 可积.由关键更新定理也可得到(34)式. □

推论 设更新过程 N 的等待时间分布 F 为非格子点分布,$\mu=\mathrm{E}[W_1]<\infty$,则$(R(t),A(t))$的极限分布为相互独立的充要条件是 N 为泊松过程.

证 记 $H(x)=\dfrac{1}{\mu}\displaystyle\int_x^\infty \overline{F}(t)\mathrm{d}t$.由于

$$\lim_{t\to\infty}\mathrm{P}(R(t)>x,A(t)\geqslant y)=\lim_{t\to\infty}\mathrm{P}(A(t+x)\geqslant x+y)=H(x+y),\quad x,y\geqslant 0,$$
$$(R(t),A(t))\text{渐近独立}\Leftrightarrow H(x+y)=H(x)H(y)\Leftrightarrow H(x)=\mathrm{e}^{-\alpha x}$$
$$\Leftrightarrow \overline{F}(x)=\mathrm{e}^{-\alpha x}\Leftrightarrow N\text{ 为泊松过程}.$$ □

定理 3.5 设延迟更新过程 N 满足条件:$\mathrm{E}[W_1]<\infty$,$\mu=\mathrm{E}[W_2]<\infty$,$\sigma^2=\mathrm{Var}[W_2]<\infty$,等待时间分布 F 为非格子点分布,则

$$\lim_{t\to\infty}\mathrm{E}[A(t)]=\frac{\mu^2+\sigma^2}{2\mu},\tag{35}$$

$$\lim_{t\to\infty}\mathrm{E}[R(t)]=\frac{\mu^2+\sigma^2}{2\mu},\tag{36}$$

$$\lim_{t\to\infty} \mathrm{E}[\beta(t)] = \frac{\mu^2+\sigma^2}{\mu}. \tag{37}$$

证 我们有

$$\mathrm{E}[R(t)] = \int_t^\infty \overline{G}(s)\mathrm{d}s + \int_0^t \int_{t-s}^\infty \overline{F}(u)\mathrm{d}u\mathrm{d}m(s),$$

$\int_t^\infty \overline{F}(s)\mathrm{d}s = \int_t^\infty (s-t)\mathrm{d}F(s)$为非负单调下降函数,且

$$\int_0^\infty \left(\int_t^\infty (s-t)\mathrm{d}F(s)\right)\mathrm{d}t = \int_0^\infty \left(\int_0^s (s-t)\mathrm{d}t\right)\mathrm{d}F(s) = \frac{1}{2}\int_0^\infty s^2\mathrm{d}F(s) = \frac{1}{2}(\mu^2+\sigma^2),$$

由引理 3.5, $\int_t^\infty \overline{F}(s)\mathrm{d}s$ 直接 R 可积. 利用关键更新定理即得(36).

我们有

$$\mathrm{E}[\beta(t)] = \int_t^\infty u\mathrm{d}G(u) + \int_0^t \int_{t-s}^\infty u\mathrm{d}F(u)\mathrm{d}m(s),$$

$\int_t^\infty u\mathrm{d}F(u)$是非负单调下降函数,且

$$\int_0^\infty \left(\int_t^\infty s\mathrm{d}F(s)\right)\mathrm{d}t = \int_0^\infty s\left(\int_0^s \mathrm{d}t\right)\mathrm{d}F(s) = \int_0^\infty s^2\mathrm{d}F(s) = (\mu^2+\sigma^2),$$

由引理 3.5, $\int_t^\infty u\mathrm{d}F(u)$直接 R 可积. 利用关键更新定理即得(37).

(35)由(36)及(37)即得. □

我们看到, $A(t)$与 $R(t)$的极限分布与极限均值都是相同的. 更有意思的是 $\beta(t) = W_{N(t)+1}$,即包含时刻 t 的等待时间的极限分布与等待时间分布 F 相去甚远,其均值的极限$(\sigma^2+\mu^2)/\mu \geqslant \mu$,只有在 $\sigma^2=0$,即等待时间是常数 μ 时才会一样,一般情况下都要变大. 这一现象在讨论泊松过程时我们已经观察到了,对更新过程它也同样存在,其道理也是相同的.

定理 3.6 设延迟更新过程 N 满足条件: $\mu_1 = \mathrm{E}[W_1] < \infty$, $\mu = \mathrm{E}[W_2] < \infty$, $\sigma^2 = \mathrm{Var}[W_2] < \infty$,等待时间分布 F 为非格子点分布,则

$$\lim_{t\to\infty}\left(m(t) - \frac{t}{\mu}\right) = \frac{\sigma^2+\mu^2}{2\mu^2} - \frac{\mu_1}{\mu}. \tag{38}$$

证 由于

$$\mathrm{E}[R(t)] = \mathrm{E}[S_{N(t)+1} - t] = \mu_1 + \mu m(t) - t,$$

所以

$$m(t) - \frac{t}{\mu} = \frac{1}{\mu}\mathrm{E}[R(t)] - \frac{\mu_1}{\mu}.$$

由(36)即得(38)式. □

定义 设一个系统依次处于"接通"与"断开"两种状态.若 X_1 为系统开始保持接通的时间,而后系统处于断开状态的时间为 Y_1,之后系统保持接通的时间为 X_2,再保持断开的时间为 Y_2 等等.记

$$W_n = X_n + Y_n, \qquad S_n = W_1 + \cdots + W_n, \quad n \geqslant 1, \quad S_0 = 0.$$

令

$$V(t) = \begin{cases} 1, & S_{n-1} \leqslant t < S_{n-1} + X_n, \\ 0, & S_{n-1} + X_n \leqslant t < S_n, \end{cases} \quad n \geqslant 1, \tag{39}$$

即 $\{V(t)=1\}$ 代表系统处于"接通"状态,$\{V(t)=0\}$ 代表系统处于"断开"状态.若 $\{(X_n, Y_n), n \geqslant 1\}$ 为 i.i.d.随机向量序列,$\{V(t), t \geqslant 0\}$ 称为交替更新过程.

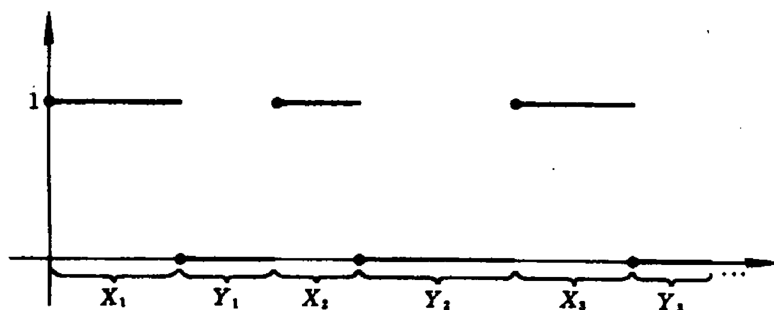


图 4

定理 3.7 对交替更新过程 $\{V(t), t \geqslant 0\}$,设 $\mathrm{E}[X_1 + Y_1] < \infty$,$W_1 = X_1 + Y_1$ 为非格子点分布的.令

$$B(t) = \mathrm{P}(\text{在时刻 } t \text{ 系统处于接通状态}),$$

则

$$\lim_{t \to \infty} B(t) = \frac{\mathrm{E}[X_1]}{\mathrm{E}[X_1] + \mathrm{E}[Y_1]}. \tag{40}$$

证 记 $\{W_i = X_i + Y_i, i \geqslant 1\}$.因为从时刻 $W_1(W_1 + W_2, W_1 + W_2 + W_3, \cdots)$ 起,交替更新过程 $\{V(t), t \geqslant 0\}$ 又重新开始,所以 $W_1, W_1 + W_2, W_1 + W_2 + W_3, \cdots$ 都是更新时刻.因此,我们可以使用更新技巧:对更新时刻 W_1 取条件期望.

$$B(t) = \int_0^\infty \mathrm{P}[V(t) = 1 \mid W_1 = s] \mathrm{d}F(s),$$

$$\mathrm{P}[V(t) = 1 \mid W_1 = s] = \begin{cases} \mathrm{P}[X_1 > t \mid W_1 = s], & s > t, \\ B(t-s), & s \leqslant t, \end{cases}$$

因为若 $s > t$,在 $W_1 = s$ 的条件下,$V(t) = 1$ 就是 $X_1 > t$;若 $s \leqslant t$,将时间原点移至 s,过程从头开始.因此

$$B(t) = \int_0^t B(t-s) \mathrm{d}F(s) + \int_t^\infty \mathrm{P}[X_1 > t \mid W_1 = s] \mathrm{d}F(s)$$

$$=\mathrm{P}(X_1>t)+\int_0^t B(t-s)\mathrm{d}F(s).$$

由引理 3.5,$\mathrm{P}(X_1>t)$是直接 R 可积函数,且

$$\int_0^\infty \mathrm{P}(X_1>t)\mathrm{d}t=\mathrm{E}[X_1],$$

再由定理 3.3 的推论,

$$\lim_{t\to\infty}B(t)=\frac{\int_0^\infty \mathrm{P}(X_1>t)\mathrm{d}t}{\mathrm{E}[W_1]}=\frac{\mathrm{E}[X_1]}{\mathrm{E}[X_1]+\mathrm{E}[Y_1]},$$

即(40)成立. □

(40)式有简单明了的直观解释,把"接通—断开"看作一个周期,那么(40)式的右端表示平均地每个周期中,系统处于接通的时间所占的比例,(40)式表明这一比例就是每个时刻系统正常工作的概率的稳定值.交替更新过程本身就是一类应用广泛的过程模型,只要一个系统重复地交替处于两个不同的状态,往往可以用交替更新过程作为模型.例如某个系统的某一元件经常失效,在它失效后,系统即停止工作,调换新的元件并进行维修,维修结束后系统才重新开始工作.若 X_i 表示第 i 个元件的寿命,Y_i 表示第 i 个元件失效后系统维修的时间.设 $\{(X_i,Y_i),i\geqslant1\}$ i.i.d..这时,系统在工作与接通相对应,系统在维修与断开相对应.当 X_1+Y_1 为非格子点分布时,长时间后系统处于工作状态的概率为$\dfrac{\mathrm{E}[X_1]}{\mathrm{E}[X_1]+\mathrm{E}[Y_1]}$.

例 3.1 对更新过程 N,$x>0$,令

$$X_n=\min(W_n,x),\quad Y_n=W_n-X_n=(W_n-x)^+,\qquad n\geqslant1,$$

则$\{(X_n,Y_n)\}$为 i.i.d.随机向量序列,可定义相应的交替更新过程$\{V(t),t\geqslant0\}$,且

$$\{A(t)\geqslant x\}=\{V(t)=0\}.$$

若等待时间分布 F 为非格子点分布,则由定理 3.7,

$$\lim_{t\to\infty}\mathrm{P}(A(t)\geqslant x)=\lim_{t\to\infty}\mathrm{P}(V(t)=0)=\frac{\mathrm{E}[Y_1]}{\mu}.$$

但$\mathrm{E}[Y_1]=\mu-\mathrm{E}[X_1]$,

$$\mathrm{E}[X_1]=\int_0^x s\mathrm{d}F(s)+\int_x^\infty x\mathrm{d}F(s)=\int_0^x \bar{F}(s)\mathrm{d}s,$$

由此即得(32)式.

类似地,令

$$X_n=(W_n-x)^+,\quad Y_n=\min(W_n,x),\qquad n\geqslant1,$$

则$\{(X_n,Y_n)\}$为 i.i.d.随机向量序列,又可定义相应的交替更新过程$\{V(t),t\geqslant0\}$,且

$$\{R(t)>x\}=\{V(t)=1\}.$$

若等待时间分布 F 为非格子点分布，则由定理 3.7，

$$\lim_{t\to\infty}\mathrm{P}(R(t)>x)=\lim_{t\to\infty}\mathrm{P}(V(t)=1)=\frac{\mathrm{E}[X_1]}{\mu}.$$

但 $\mathrm{E}[X_1]=\mu-\mathrm{E}[Y_1]$，$\mathrm{E}[Y_1]=\int_0^x\overline{F}(s)\mathrm{d}s$，由此即得(33)式.

本例提供的技巧把讨论的对象化为交替更新过程，在解许多问题时非常有用. □

例 3.2 ***M/G/1* 消失系统** 设顾客按照参数为 λ 的泊松过程来到某个只有一个理发师的理发店. 每个顾客到来后，若已有顾客在理发，则立即离开. 若理发店无顾客，则新来的顾客就进门理发. 假定理发的时间是一个随机变量，其分布为 G(与到来的顾客无关). 现在问，一个顾客来到理发店可马上开始理发的概率是多少？

解 我们把理发店看为一个系统，它依次交替处于"空闲"和"忙碌"两个状态. 由于指数分布的无记忆性，系统空闲的时间间隔 $\{X_k,k\geqslant1\}$ 为 i.i.d. 随机变量序列，X_k 为指数分布的，$\mathrm{E}[X_k]=1/\lambda$，系统忙碌的时间间隔 $\{Y_k,k\geqslant1\}$ 也为 i.i.d. 随机变量序列，且与 $\{X_k\}$ 独立.

$$\mathrm{E}[Y_k]=\int_0^\infty y\mathrm{d}G(y)=m_G.$$

若 $p(t)=\mathrm{P}$(在时刻 t 系统处于空闲)，则由(40)

$$\lim_{t\to\infty}p(t)=\frac{\mathrm{E}[X_1]}{\mathrm{E}[X_1]+\mathrm{E}[Y_1]}=\frac{1/\lambda}{1/\lambda+m_G}=\frac{1}{1+\lambda m_G}.$$

这便是长时间后，一个顾客到达理发店时可马上开始理发的概率.

特别地，若 $\lambda=2$(平均每小时两个顾客到来)，$m_G=1/3$，那么 $\lim_{t\to\infty}p(t)=3/5$，即只有 3/5 的顾客被服务. □

例 3.3 ***M/G/1* 排队系统** 现在我们讨论正常的 $M/G/1$ 排队系统，即在上例中，到来的顾客发现理发店已有顾客在理发时，不马上走开，而是排队等候. 我们仍然求一个顾客来到理发店可马上开始理发的概率.

解 这时系统仍交替地处于空闲与忙碌两种状态，不过现在理发师要把到来顾客(包括在服务过程中到来的顾客)都服务完，系统才空闲，等待下一个顾客到来. 仍由于指数分布的无记忆性，空闲的时间间隔 $\{X_k,k\geqslant1\}$ 为 i.i.d. 随机变量序列，且 $\mathrm{E}[X_k]=1/\lambda$. 而忙碌时段 $\{Y_k,k\geqslant1\}$ 亦为 i.i.d. 随机变量序列，且与 $\{X_k\}$ 独立，所以对 $p(t)=\mathrm{P}$(时刻 t 系统空闲)有

$$\lim_{t\to\infty}p(t)=\frac{\mathrm{E}[X_1]}{\mathrm{E}[X_1]+\mathrm{E}[Y_1]}=\frac{1}{1+\lambda\ \mathrm{E}[Y_1]}.\tag{41}$$

为了计算 $\mathrm{E}[Y_1]$，以 U 表示第一个顾客服务过程中新来到的顾客数，V 表示第一个顾客的服务时间，则已知 U 与 V 时

$$\mathrm{E}[Y_1\mid U=u,V=v]=v+u\ \mathrm{E}[Y_1].$$

所以

$$\mathrm{E}[Y_1]=\mathrm{E}[V+U\ \mathrm{E}[Y_1]]=\mathrm{E}[V]+\mathrm{E}[U]\mathrm{E}[Y_1].\tag{42}$$

为了保证理发店的排队人数不致无限增长，来往强度 λ/μ 应小于1，这里 $\mu^{-1}=\mathrm{E}[V]$，即平均地说，一个顾客的服务时间内，新来的顾客数应该小于1，因为

$$\mathrm{E}[U]=\int \mathrm{E}[U\mid V=v]\mathrm{d}H(v)=\int(\lambda v)\mathrm{d}H(v)=\frac{\lambda}{\mu},$$

其中 H 是 V 的分布函数. 所以由(42)

$$\mathrm{E}[Y_1]=\frac{\mathrm{E}[V]}{1-\mathrm{E}[U]}=\frac{1/\mu}{1-\lambda/\mu}=\frac{1}{\mu-\lambda}.$$

又由(41)

$$\lim_{t\to\infty}p(t)=\frac{1}{1+\lambda/(\mu-\lambda)}=1-\frac{\lambda}{\mu}.$$

特别地，对 $\lambda=2,\mu=3,\lim_{t\to\infty}p(t)=1/3$. 与例 3.2 相比较，消失系统拒绝了 2/5 的顾客，但接受服务的顾客都不需等待. 而正常的排队系统为每个到来的顾客服务，但有 2/3 的顾客都需排队等待. 消失系统平均有 3/5 的时间是空闲着，而排队系统平均只有 1/3 时间是空闲着. □

在本节的最后，讨论格子点分布情形的布莱克威尔定理及关键更新定理. 为叙述简单起见，只讨论更新过程. 延迟更新过程留给读者去讨论.

定理 3.8 设更新过程 N 的等待时间分布是步长为 d 的格子点分布，则

$$\lim_{n\to\infty}[m(nd)-m((n-1)d)]=d/\mu. \tag{43}$$

若 $0\leqslant a<d$，非负函数 h 使得 $\sum_{k=0}^{\infty}h(a+kd)<\infty$，则

$$\lim_{n\to\infty}\int_0^{a+nd}h(a+nd-s)\mathrm{d}m(s)=\frac{d}{\mu}\sum_{k=0}^{\infty}h(a+kd). \tag{44}$$

证 令 $a_n=\mathrm{P}(W_1=nd),n\geqslant0$，则 $\{n\geqslant1:a_n>0\}$ 的最大公因子为 1，且 $\sum_{n=1}^{\infty}na_n=\mu/d$. 这时 $m(t)$ 也是格子点的，步长为 d. 令 $m_n=m(nd),n\geqslant0$，则更新函数满足的更新方程的形式为：

$$m_n=\sum_{j=0}^{n}a_j+\sum_{k=0}^{n}m_{n-k}a_k,\qquad n\geqslant0. \tag{45}$$

记

$$u_0=m_0=\frac{a_0}{1-a_0},\qquad u_n=m_n-m_{n-1},\qquad n\geqslant1.$$

由(45)，$n\geqslant1$ 时

$$\begin{aligned}u_n&=a_n+\sum_{k=0}^{n-1}m_{n-k}a_k-\sum_{k=0}^{n-1}m_{n-1-k}a_k\\&=a_n+\sum_{k=0}^{n-1}(m_{n-k}-m_{n-1-k})a_k+m_0a_n\end{aligned}$$

$$= a_n + \sum_{k=0}^{n} u_{n-k} a_k$$

$$= a_n + a_0 u_n + \sum_{k=1}^{n} a_k u_{n-k}. \tag{46}$$

记

$$f_n = \begin{cases} \dfrac{a_n}{1-a_0}, & n \geqslant 1, \\ 0, & n = 0, \end{cases}$$

则

$$\sum_{n=1}^{\infty} f_n = 1, \qquad \sum_{n=1}^{\infty} n f_n = \frac{\mu}{d(1-a_0)}.$$

由(46)得

$$u_n = f_n + \sum_{k=1}^{n} f_k u_{n-k}, \qquad n \geqslant 1. \tag{47}$$

记 $U(z) = \sum_{n=0}^{\infty} u_n z^n, F(z) = \sum_{n=0}^{\infty} f_n z^n$,则(47)可写成

$$U(z) - u_0 = F(z) + F(z)U(z),$$

$$U(z) = \frac{u_0 + F(z)}{1 - F(z)}. \tag{48}$$

令 $P(z) = \sum_{n=0}^{\infty} p_n z^n = 1/[1-F(z)]$,则 $P(z) = F(z)P(z) + 1$,即

$$p_0 = 1, \qquad p_n = \sum_{k=1}^{n} f_k p_{n-k}, \qquad n \geqslant 1.$$

由马尔可夫链的转移概率的极限性质(第二章定理 6.1)可知

$$\lim_{n \to \infty} p_n = \frac{d}{\mu}(1 - a_0).$$

由(48), $U(z) = [u_0 + F(z)]P(z)$,故(参见第二章引理 3.1)

$$u_n = u_0 p_n + \sum_{k=1}^{n} f_k p_{n-k} \to \left(u_0 + \sum_{k=1}^{\infty} f_k\right)\frac{d}{\mu}(1 - a_0) = \frac{d}{\mu},$$

(43)得证.(44)可由下式类似地得到:

$$\int_0^{a+nd} h(a + nd - s)\mathrm{d}m(s) = \sum_{k=0}^{n} h(a + (n-k)d) u_k. \qquad \square$$

习　　题

3-1. 试由布莱克威尔定理推出初等更新定理.

3-2. 设 F 非格子点分布,证明:

$$\lim_{t\to\infty}[m(t)-m(t^-)]=0.$$

3－3. 设延迟更新过程 N 的更新函数 $m(t)$ 为线性函数：$m(t)=Ct, C>0$ 为常数，则 N 为稳定更新过程.

3－4. 设更新过程 $\{N(t),t\geqslant 0\}$ 满足条件 $\mathsf{E}[W_1]=\mu<\infty$， $\mathsf{Var}[W_1]=\sigma^2<\infty$. 证明：

$$\lim_{t\to\infty}\frac{\mathsf{Var}[N(t)]}{t}=\frac{\sigma^2}{\mu^3}.$$

3－5. 证明：(1) 若 $h\geqslant 0$ 在 $[0,\infty)$ 上直接 R 可积，则 $h*F$ 亦然.

(2) 若 h_1,h_2 为 $[0,\infty)$ 上直接 R 可积的非负函数，则 h_1+h_2 亦然.

3－6. 假设延迟更新过程 N 的等待时间分布 F 为非格子点分布. 证明下列命题等价：

(1) 布莱克威尔定理；

(2) 关键更新定理；

(3) $\lim_{t\to\infty}\mathsf{P}(A(t)\geqslant x)=\frac{1}{\mu_F}\int_x^\infty\overline{F}(t)\mathrm{d}t,\quad x\geqslant 0$;

(4) $\lim_{t\to\infty}\mathsf{P}(R(t)>x)=\frac{1}{\mu_F}\int_x^\infty\overline{F}(t)\mathrm{d}t,\quad x\geqslant 0$.

3－7. 设更新过程 N 的等待时间分布为非格子点分布，证明：

1) 若 $p>0$，且 $\mathsf{E}[W_1^{p+1}]<\infty$，则 $\mathsf{E}[A(t)^p]$，$\mathsf{E}[R(t)^p]$，$\mathsf{E}[\beta(t)^p]$ 存在且

$$\lim_{t\to\infty}\mathsf{E}[A(t)^p]=\lim_{t\to\infty}\mathsf{E}[R(t)^p]=\frac{\mathsf{E}[W_1^{p+1}]}{(p+1)\mathsf{E}[W_1]},$$

$$\lim_{t\to\infty}\mathsf{E}[\beta(t)^p]=\frac{\mathsf{E}[W_1^{p+1}]}{\mathsf{E}[W_1]},$$

2) 若 $\mathsf{E}[W_1^3]<\infty$，则

$$\lim_{t\to\infty}\mathsf{E}[A(t)R(t)]=\frac{\mathsf{E}[W_1^3]}{6\ \mathsf{E}[W_1]}.$$

3－8. 设在Ⅱ型计数器中，信号来到的间隔服从非格子点分布 F. 每个来到的信号使计数器关闭一段固定长度 d 的时间. 记 $B(t)$ 为时刻 t 计数器不关闭的概率. 若 $\mu=\int_0^\infty\overline{F}(t)\mathrm{d}t<\infty$，证明：

$$\lim_{t\to\infty}B(t)=\frac{1}{\mu}\int_d^\infty\overline{F}(t)\mathrm{d}t.$$

3－9. 某工作台上的仪器需日夜运行. 当仪器损坏时便换一台新的仪器. 试问在下列情况下，长时间后去检查在使用的仪器，其使用寿命（以年为单位）不到一年的概率是多少？

1) 每台仪器的寿命分布为 $(0,2)$ 上的均匀分布；

2) 每台仪器的寿命分布是均值为 1 的指数分布.

3－10. 在Ⅰ型计数器（例 1.1）中，设信号来到间隔的分布为 F，信号使计数器关闭的时间的分布为 G，F,G 为非格子点分布，$\mu_F<\infty$，$\mu_G<\infty$. 以 $B(t)$ 记时刻 t 计数器关闭的概率，证明：

$$\lim_{t\to\infty}B(t)=\frac{\mu_G}{\mu_F[1+\int_0^\infty m_F(t)\mathrm{d}G(t)]}.$$

3－11. 设$\{(X_n^{(1)},\cdots,X_n^{(d)}),n\geqslant1\}$为 i.i.d. d 维随机向量序列. 记

$$W_n=X_n^{(1)}+\cdots+X_n^{(d)},\quad S_n=W_1+\cdots+W_n,\quad n\geqslant1,\quad S_0=0.$$

令

$$V(t)=\begin{cases}0, & S_{n-1}\leqslant t<S_{n-1}+X_n^{(1)},\\ j-1, & S_{n-1}+X_n^{(1)}+\cdots+X_n^{(j-1)}\leqslant t<S_{n-1}+X_n^{(1)}+\cdots+X_n^{(j)},\end{cases}$$
$$j=2,\cdots,d,\quad n\geqslant1.$$

$\{V(t),t\geqslant0\}$可称为多状态交替更新过程. 若 W_1 为非格子点分布的, 则

$$\lim_{t\to\infty}\mathrm{P}(V(t)=j-1)=\frac{\mathrm{E}[W_1^{(j)}]}{\mathrm{E}[X_1^{(1)}+\cdots+X_1^{(d)}]},\quad j=1,\cdots,d.$$

3－12. 对交替更新过程 N, 定义

$$A_1(t)=\begin{cases}t-W_{n-1}, & W_{n-1}\leqslant t<W_{n-1}+X_n,\quad n\geqslant1,\\ 0, & \text{其它},\end{cases}$$
$$R_1(t)=\begin{cases}W_{n-1}+X_n-t, & W_{n-1}\leqslant t<W_{n-1}+X_n,\quad n\geqslant1,\\ 0, & \text{其它},\end{cases}$$
$$A_2(t)=\begin{cases}t-(W_{n-1}+X_n), & W_{n-1}+X_n\leqslant t<W_n,\quad n\geqslant1,\\ 0, & \text{其它},\end{cases}$$
$$R_2(t)=\begin{cases}W_n-t, & W_{n-1}+X_n\leqslant t<W_n,\quad n\geqslant1,\\ 0 & \text{其它},\end{cases}$$

设 W_1 为非格子点分布的. 对 $x>0$ 求:

$$\lim_{t\to\infty}\mathrm{P}(A_i(t)\geqslant x),\quad \lim_{t\to\infty}\mathrm{P}(R_i(t)>x),\quad i=1,2.$$

§4.4 有偿更新过程与可终止更新过程

定义 设$\{(W_n,Y_n),n\geqslant1\}$为 i.i.d. 二维随机变量序列, $W_n\geqslant0$, N 是以$\{W_n,\geqslant1\}$为等待时间的更新过程, 令

$$X(t)=\sum_{n=1}^{N(t)}Y_n,$$

则 $X=\{X(t),t\geqslant0\}$称为有偿更新过程. 当第 n 次更新发生时, 就获得奖赏或惩罚 Y_n. $X(t)$就是到时刻 t 为止累积的奖赏或惩罚. 值得指出的是 X_n 与 Y_n 并不需要假定相互独立, 也不必假设 $Y_n\geqslant0$. 但是

$$X(t)=X_1(t)-X_2(t), X_1(t)=\sum_{n=1}^{N(t)} Y_n^+, X_2(t)=\sum_{n=1}^{N(t)} Y_n^-,$$

$X_1(t), X_2(t)$都是有偿更新过程,因此往往只需讨论奖赏非负的有偿更新过程.

定理 4.1 设 X 为有偿更新过程, $\mathrm{E}[W_1]<\infty$, $\mathrm{E}[|Y_1|]<\infty$,则以概率 1

$$\lim_{t\to\infty}\frac{X(t)}{t}=\frac{\mathrm{E}[Y_1]}{\mathrm{E}[W_1]}, \tag{1}$$

$$\lim_{t\to\infty}\frac{\mathrm{E}[X(t)]}{t}=\frac{\mathrm{E}[Y_1]}{\mathrm{E}[W_1]}. \tag{2}$$

证 由强大数律及定理 1.1 即得(1):

$$\frac{X(t)}{t}=\frac{\sum_{i=1}^{N(t)} Y_i}{N(t)}\cdot\frac{N(t)}{t}\to\frac{\mathrm{E}[Y_1]}{\mathrm{E}[W_1]}.$$

由瓦尔德等式

$$\mathrm{E}[X(t)]=\mathrm{E}\left[\sum_{i=1}^{N(t)+1} Y_i\right]-\mathrm{E}[Y_{N(t)+1}]=\mathrm{E}[N(t)+1]\mathrm{E}[Y_1]-\mathrm{E}[Y_{N(t)+1}].$$

所以只需证明

$$\lim_{t\to\infty}\frac{1}{t}\mathrm{E}[Y_{N(t)+1}]=0, \tag{3}$$

即可由初等更新定理推出(2).为证(3),令 $B(t)=\mathrm{E}[Y_{N(t)+1}]$,并用"更新技巧":

$$\mathrm{E}[Y_{N(t)+1}\mid W_1=s]=\begin{cases}\mathrm{E}[Y_1\mid W_1=s], & s>t,\\ B(t-s), & s\leqslant t,\end{cases}$$

因为 $W_1=s>t$ 时, $N(t)=0$,而 $s\leqslant t$ 时,将时间原点移至 s,过程重新开始.所以

$$B(t)=\int_0^\infty \mathrm{E}[Y_{N(t)+1}\mid W_1=s]\mathrm{d}F(s)=b(t)+\int_0^t B(t-s)\mathrm{d}F(s),$$

其中 $b(t)=\int_t^\infty \mathrm{E}[Y_1\mid W_1=s]\mathrm{d}F(s)$.事实上,对一切 t,

$$\begin{aligned}|b(t)|&\leqslant\int_t^\infty |\mathrm{E}[Y_1\mid W_1=s]|\mathrm{d}F(s)\leqslant\int_t^\infty \mathrm{E}[|Y_1|\mid W_1=s]\mathrm{d}F(s)\\&\leqslant\int_0^\infty \mathrm{E}[|Y_1|\mid W_1=s]\mathrm{d}F(s)=\mathrm{E}[|Y_1|]<\infty.\end{aligned} \tag{4}$$

因此有

$$B(t)=b(t)+\int_0^t b(t-s)\mathrm{d}m(s).$$

由(4), $t\to\infty$ 时, $b(t)\to 0$,故对任给 $\varepsilon>0$,有 $T>0$,使得 $t>T$ 时, $|b(t)|<\varepsilon$,从而

$$\frac{|B(t)|}{t}\leqslant\frac{|b(t)|}{t}+\left(\int_0^{t-T}+\int_{t-T}^{t}\right)\frac{|b(t-s)|}{t}\mathrm{d}m(s)$$
$$\leqslant\frac{\varepsilon}{t}+\varepsilon\frac{m(t-T)}{t}+\mathsf{E}[|Y_1|]\frac{m(t)-m(t-T)}{t}.$$

在上式中令 $t\to\infty$,由初等更新定理得

$$\overline{\lim_{t\to\infty}}\frac{|B(t)|}{t}\leqslant\frac{\varepsilon}{\mathsf{E}[W_1]},$$

再令 $\varepsilon\to0$ 即得(3). □

(2) 表明对有偿更新过程,其长时间内获得的平均奖赏近似于每个更新周期内获得的单位时间内的平均奖赏.这在实际中有广泛的应用.

例 4.1 设某人长期使用自行车,一辆车的寿命为 Z 个月,Z 的分布函数为 H.若原有的车辆用了 T(T 为一固定正数)个月尚未损坏,为安全起见将它弃之不用,化 c_1 元换一辆新车.若在 T 个月以前损坏,除化费 c_1 元换新车外还需附加 c_2 元的其它支出.问此人平均每月要支付多少费用?

解 我们用有偿更新过程来描述这一问题.若第 n 辆车可使用 Z_n 个月才损坏,但实际使用期为 $W_n=\min(Z_n,T)$个月,$n\geqslant1$.为换第 n 辆新车所需的化费为 Y_n,按题设

$$Y_n=c_1 1_{\{W_n\geqslant T\}}+(c_1+c_2)1_{\{W_n<T\}}.$$

因而由定理 4.1,长时间使用自行车的平均化费是

$$\lim_{t\to\infty}\frac{X(t)}{t}=\frac{\mathsf{E}[Y_1]}{\mathsf{E}[W_1]}=\frac{c_1+c_2\mathsf{P}(W_1<T)}{\mathsf{E}[W_1]}=\frac{c_1+c_2H(T^-)}{\int_0^T\overline{H}(x)\mathrm{d}x}.$$

为了使平均化费最小,应选取 T,使上式右端达到最小. □

定理 4.2 设 X 为有偿更新过程,W_1 为非格子点变量,$\mathsf{E}[W_1]<\infty$,$\mathsf{E}[Y_1]<\infty$,$\mathsf{E}[W_1Y_1]<\infty$,则对任意的 $a>0$,

$$\lim_{t\to\infty}(\mathsf{E}[X(t+a)]-\mathsf{E}[X(t)])=a\frac{\mathsf{E}[Y_1]}{\mathsf{E}[W_1]}.\tag{5}$$

证 由瓦尔德等式,

$$\mathsf{E}[X(t+a)]-\mathsf{E}[X(t)]=\mathsf{E}\Big[\sum_{i=1}^{N(t+a)+1}Y_i\Big]-\mathsf{E}[Y_{N(t+a)+1}]-\mathsf{E}\Big[\sum_{i=1}^{N(t)+1}Y_i\Big]+\mathsf{E}[Y_{N(t)+1}]$$
$$=(\mathsf{E}[N(t+a)]-\mathsf{E}[N(t)])\mathsf{E}[Y_1]-\mathsf{E}[Y_{N(t+a)+1}]+\mathsf{E}[Y_{N(t)+1}].$$

所以只需证明

$$\lim_{t\to\infty}\mathsf{E}[Y_{N(t)+1}]\text{存在且有限},\tag{6}$$

即可由布莱克威尔定理推出(5).为证(6),令 $B(t)=\mathsf{E}[Y_{N(t)+1}]$.在定理 4.1 的证明中我们已经知道

$$B(t)=b(t)+\int_0^t b(t-s)\mathrm{d}m(s),$$

其中 $b(t)=\int_t^\infty \mathrm{E}[Y_1|W_1=s]\mathrm{d}F(s)$. 不妨设 $Y_1\geqslant 0$,则 $b(t)=\int_t^\infty \mathrm{E}[Y_1|W_1=s]\mathrm{d}F(s)$是单调下降的非负函数,且

$$\int_0^\infty b(t)\mathrm{d}t=\int_0^\infty\left(\int_t^\infty \mathrm{E}[Y_1|W_1=s]\mathrm{d}F(s)\right)\mathrm{d}t=\int_0^\infty s\,\mathrm{E}[Y_1|W_1=s]\mathrm{d}F(s)$$
$$=\int_0^\infty \mathrm{E}[W_1Y_1|W_1=s]\mathrm{d}F(s)=\mathrm{E}[W_1Y_1]<\infty,$$

因此 $b(t)$是直接 R 可积的,这样由关键更新定理就得到(6):

$$\lim_{t\to\infty}B(t)=\frac{\mathrm{E}[W_1Y_1]}{\mathrm{E}[W_1]}.\qquad\square$$

定理 4.3 设 N 为更新过程,$\mu=\mathrm{E}[W_1]<\infty,\sigma^2=\mathrm{Var}[W_1]<\infty$,则以概率 1

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t A(s)\mathrm{d}s=\frac{\mu^2+\sigma^2}{2\mu},\tag{7}$$

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t R(s)\mathrm{d}s=\frac{\mu^2+\sigma^2}{2\mu},\tag{8}$$

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \mathrm{E}[A(s)]\mathrm{d}s=\frac{\mu^2+\sigma^2}{2\mu},\tag{9}$$

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \mathrm{E}[R(s)]\mathrm{d}s=\frac{\mu^2+\sigma^2}{2\mu}.\tag{10}$$

证 令

$$Y_n=\int_{S_{n-1}}^{S_n}A(s)\mathrm{d}s=\int_0^{W_n}A(s+S_{n-1})\mathrm{d}s=\int_0^{W_n}s\mathrm{d}s=\frac{W_n^2}{2}.$$

由此易见$\{(W_n,Y_n),n\geqslant 1\}$为 i.i.d. 的. 以$\{X(t),t\geqslant 0\}$表示相应的有偿更新过程,则

$$X(t)=\sum_{n=1}^{N(t)}Y_n\leqslant\int_0^t A(s)\mathrm{d}s\leqslant\sum_{n=1}^{N(t)+1}Y_n,\tag{11}$$

$$\mathrm{E}[X(t)]\leqslant\int_0^t \mathrm{E}[A(s)]\mathrm{d}s\leqslant\mathrm{E}[Y_1][1+m(t)].\tag{12}$$

注意到$\mathrm{E}[Y_1]=\mathrm{E}[W_1^2/2]=(\sigma^2+\mu^2)/2$,利用强大数律及定理 4.1,由(11)及(12)即得(7)及(9).

另一方面,又有

$$\int_{S_{n-1}}^{S_n}R(s)\mathrm{d}s=\int_0^{W_n}R(s+S_{n-1})\mathrm{d}s=\int_0^{W_n}(W_n-s)\mathrm{d}s=\frac{W_n^2}{2}=Y_n.$$

因此,同样地可得(8)及(10). $\square$

定理 4.4 设$\{V(t),t\geqslant 0\}$为交替更新过程,$\mathrm{E}[W_1]=\mathrm{E}[X_1+Y_1]<\infty$,则以概率 1

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t 1_{\{V(s)=1\}}\mathrm{d}s=\frac{\mathsf{E}[X_1]}{\mathsf{E}[X_1+Y_1]},\tag{13}$$

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \mathsf{P}(V(s)=1)\mathrm{d}s=\frac{\mathsf{E}[X_1]}{\mathsf{E}[X_1+Y_1]}.\tag{14}$$

证 注意到

$$\int_{S_{n-1}}^{S_n} 1_{\{V(s)=1\}}\mathrm{d}s=\int_{S_{n-1}}^{S_n+X_n}\mathrm{d}s=X_n,$$

$\{(W_n,X_n),n\geqslant 1\}$ 为 i.i.d. 的，以 $\{X(t),t\geqslant 0\}$ 记相应的有偿更新过程，则

$$X(t)=\sum_{n=1}^{N(t)}X_n\leqslant\int_0^t 1_{\{V(s)=1\}}\mathrm{d}s\leqslant\sum_{n=1}^{N(t)+1}X_n,\tag{15}$$

$$\mathsf{E}[X(t)]\leqslant\int_0^t \mathsf{P}(V(s)=1)\mathrm{d}s\leqslant \mathsf{E}[X_1][1+m(t)].\tag{16}$$

利用强大数律及定理 4.1，由(15)及(16)即得(13)及(14). □

(9)及(10)是定理 3.5 中的(35)及(36)的推论，但那里需假设等待时间分布为非格子点的. 同样地，(14)是定理 3.7 的推论，那里也需要非格子点的假设.

定义 更新过程 N，若其更新时间间隔 $\{W_j,j\geqslant 1\}$ 以正概率可取无穷：

$$L=F(\infty)=\mathsf{P}(W_1<\infty)<1,\tag{17}$$

则 N 称为可终止更新过程. 这时 $N(\infty)$ 为更新总数，

$$\zeta=S_{N(\infty)}=\sup\{S_n:S_n<\infty\}$$

为最后一次更新时刻.

定理 4.5 设 N 为可终止更新过程，则更新总数 $N(\infty)$ 服从几何分布：

$$\mathsf{P}(N(\infty)=k)=(1-L)L^k,\qquad k\geqslant 0,\tag{18}$$

$$m(\infty)=\mathsf{E}[N(\infty)]=\frac{L}{1-L}.\tag{19}$$

最后更新时刻 ζ 的分布为

$$\mathsf{P}(\zeta\leqslant t)=(1-L)(1+m(t)),\tag{20}$$

$$\mathsf{E}\,\zeta=\frac{1}{1-L}\int_0^\infty (L-F(s))\mathrm{d}s.\tag{21}$$

证 由于

$$\{N(\infty)=k\}=\{W_1<\infty,\cdots,W_k<\infty,W_{k+1}=\infty\},$$
$$=\mathsf{P}(W_1<\infty)\cdots\mathsf{P}(W_k<\infty)\mathsf{P}(W_{k+1}=\infty),\quad k\geqslant 0,$$

故得(16)式. 所以

$$m(\infty)=\mathrm{E}[N(\infty)]=\sum_{k=1}^{\infty}\mathrm{P}(N(\infty)\geqslant k)=\sum_{k=1}^{\infty}L^k=\frac{L}{1-L}.$$

$$\mathrm{P}(\zeta\leqslant t)=\sum_{k=0}^{\infty}\mathrm{P}(\zeta\leqslant t,N(\infty)=k)=\sum_{k=0}^{\infty}\mathrm{P}(S_k\leqslant t,W_{k+1}=\infty)$$

$$=\sum_{k=0}^{\infty}\mathrm{P}(S_k\leqslant t)(1-L)=(1+m(t))(1-L).$$

$$\mathrm{E}\,\zeta=\mathrm{E}[\zeta 1_{(W_1<\infty)}]=\int_0^{\infty}\mathrm{E}[\zeta\mid W_1=s]\mathrm{d}F(s)$$

$$=\int_0^{\infty}(s+\mathrm{E}[\zeta])\mathrm{d}F(s)=\int_0^{\infty}s\mathrm{d}F(s)+L\,\mathrm{E}[\zeta],$$

所以有

$$\mathrm{E}\,\zeta=\frac{1}{1-L}\int_0^{\infty}s\mathrm{d}F(s)=\frac{1}{1-L}\int_0^{\infty}(L-F(s))\mathrm{d}s. \qquad \square$$

例 4.2 设某一条马路上通过的车辆数是一个更新过程,其更新间隔$\{W_i,i\geqslant 1\}$的公共分布为 F. 假设一个行人欲穿过这一马路,必须是前后车辆间隔时间超过 τ 时才能通过道路,于是这个行人在更新时刻 S_n 首次可以穿过马路当且仅当 $W_1\leqslant\tau,\cdots,W_n\leqslant\tau,W_{n+1}>\tau$. 我们来求首次可穿过马路的时刻的分布与期望. 令

$$\widetilde{W}_i=W_i 1_{\{W_i\leqslant\tau\}}+(+\infty)1_{\{W_i>\tau\}},$$

则 $\widetilde{W}_i$ 的分布为

$$\widetilde{F}(t)=F(\min(t,\tau)),$$

而首次可穿过马路的时刻便是以$\{\widetilde{W}_i,i\geqslant 1\}$为更新间隔的可终止更新过程 $\widetilde{N}$ 的最后更新时刻 $\widetilde{\zeta}$,故由定理 4.5

$$\mathrm{P}(\zeta\leqslant t)=(1-\widetilde{F}(\infty))(1+\widetilde{m}(t))=(1-F(\tau))(1+\widetilde{m}(t)),$$

其中 $\widetilde{m}=\mathrm{E}[\widetilde{N}(t)]$,而

$$\mathrm{E}\,\zeta=\frac{1}{1-\widetilde{F}(\infty)}\int_0^{\infty}[\widetilde{F}(\infty)-\widetilde{F}(t)]\mathrm{d}t=\frac{1}{1-F(\tau)}\int_0^{\tau}[F(\tau)-F(t)]\mathrm{d}t. \qquad \square$$

例 4.3 **Ⅱ型计数器** 假设当计数器接收到一个粒子后将封闭长为 τ 的一个时段,且若在这个时段中又有粒子到来,则从粒子到来时起还将封闭长为 τ 的一个时段,直到在长为 τ 的时段中无粒子到来,计数器才重新开放计数. 这样的计数器称为Ⅱ型计数器. 若粒子按更新过程来到,$\{W_i,i\geqslant 1\}$为粒子来到时间间隔,取

$$\widetilde{W}_i=W_i 1_{\{W_i\leqslant\tau\}}+(+\infty)1_{\{W_i>\tau\}}.$$

又记 $\widetilde{\zeta}$ 为对应于更新间隔为$\{\widetilde{W}_i,i\geqslant 1\}$的可终止更新过程的最后更新时刻,则 $\widetilde{\zeta}+\tau$ 便是一个粒子到来后计数器封闭的时间长度. 由定理 4.5 可求出它的分布与期望. $\square$

可终止更新过程 N 的更新函数 $m(t)=\mathrm{E}[N(t)]$是有界的:

$$m(t) \leqslant m(\infty) = \mathrm{E}[N(\infty)] = \frac{L}{1-L}.$$

与定理 1.3 一样,它也满足方程:

$$m(t) = F(t) + \int_0^t m(t-s)\mathrm{d}F(s).$$

设 $h(t)$为局部有界函数,这时方程

$$z(t) = h(t) + \int_0^t z(t-s)\mathrm{d}F(s), \qquad t \geqslant 0, \tag{22}$$

仍称为更新方程,尽管这时 F 满足条件(17). 与定理 1.4 一样可证,这更新方程有唯一的局部有界解:

$$z(t) = h(t) + \int_0^t h(t-s)\mathrm{d}m(s), \qquad t \geqslant 0.$$

若又有极限$\lim\limits_{t\to\infty} h(t) = h(\infty)$存在,则

$$\begin{aligned}\lim_{t\to\infty} z(t) &= \lim_{t\to\infty}\left\{h(t) + \int_0^t h(t-s)\mathrm{d}m(s)\right\} \\ &= h(\infty) + h(\infty)\int_0^\infty \mathrm{d}m(s) = h(\infty)\left[1 + \frac{L}{1-L}\right] = \frac{h(\infty)}{1-L}.\end{aligned}$$

为了将条件(17)下的更新方程(22)化为普通的更新方程,我们将假设存在正数 α 使得

$$\int_0^\infty \mathrm{e}^{\alpha t}\mathrm{d}F(t) = 1. \tag{23}$$

易见,满足(23)的正数至多只有一个. 令

$$\hat{F}(t) = \int_0^t \mathrm{e}^{\alpha s}\mathrm{d}F(s),$$

则$\hat{F}(t)$是概率分布函数:$\hat{F}(\infty) = 1$. 令

$$\hat{z}t = \mathrm{e}^{\alpha t}z(t), \qquad \hat{h}(t) = \mathrm{e}^{\alpha t}h(t),$$

则更新方程(22)化为

$$\hat{z}(t) = \mathrm{e}^{\alpha t}h(t) + \int_0^t \mathrm{e}^{\alpha(t-s)}z(t-s)\mathrm{e}^{\alpha s}\mathrm{d}F(s),$$

$$\hat{z}(t) = \hat{h}(t) + \int_0^t \hat{z}(t-s)\mathrm{d}\hat{F}(s). \tag{24}$$

这就归结为原来讨论的更新方程了.

定理 4.6 设 F 为满足(17)的非格子点分布,条件(23)满足,局部有界函数 $h(t)$有极限 $\lim\limits_{t\to\infty} h(t) = h(\infty)$,$z(t)$为更新方程(22)的解. 令

$$\hat{\mu} = \int_0^\infty t\,\mathrm{d}\hat{F}(t) = \int_0^\infty t\mathrm{e}^{\alpha t}\mathrm{d}F(t),$$

$$h_1(t)=h(t)-h(\infty)-h(\infty)\frac{L-F(t)}{1-L}.$$

若 $e^{\alpha t}h_1(t)$ 直接 R 可积,则

$$\lim_{t\to\infty}e^{\alpha t}[z(t)-z(\infty)]=\frac{1}{\hat{\upsilon}}\left\{\int_0^\infty e^{\alpha t}[h(t)-h(\infty)]dt-\frac{h(\infty)}{\alpha}\right\},\tag{25}$$

其中

$$z(\infty)=\lim_{t\to\infty}z(t)=\frac{h(\infty)}{1-L}.$$

证 若 $h(\infty)=0$,则 $h_1(t)=h(t)$,对更新方程(22)的解 $\hat{z}(t)$ 用关键更新定理即得(25)的简化形式:

$$\lim_{t\to\infty}e^{\alpha t}z(t)=\frac{1}{\hat{\mu}}\int_0^\infty e^{\alpha t}h(t)dt.$$

下面讨论 $h(\infty)\neq 0$ 的情形.

$$1=1-F(t)+\int_0^t dF(s),$$

$$z(\infty)=\frac{h(\infty)}{1-L}[1-F(t)]+\int_0^t z(\infty)dF(s),$$

$$z(t)-z(\infty)=h(t)-\frac{h(\infty)}{1-L}[1-F(t)]+\int_0^t[z(t-s)-z(\infty)]dF(s)$$

$$=h_1(t)+\int_0^t[z(t-s)-z(\infty)]dF(s),$$

$$e^{\alpha t}[z(t)-z(\infty)]=e^{\alpha t}h_1(t)+\int_0^t e^{\alpha(t-s)}[z(t-s)-z(\infty)]d\hat{F}(s).\tag{26}$$

对更新方程(24)的解用关键更新定理得

$$\lim_{t\to\infty}e^{\alpha t}[z(t)-z(\infty)]=\frac{1}{\hat{\mu}}\int_0^\infty e^{\alpha t}h_1(t)dt.$$

为得到(25)式,只需利用条件(23)及分部积分

$$\begin{aligned}\int_0^\infty e^{\alpha t}\frac{L-F(t)}{1-L}dt&=\frac{1}{\alpha}\int_0^\infty\frac{L-F(t)}{1-L}de^{\alpha t}\\&=\frac{1}{\alpha}\left\{-\frac{L}{1-L}+\frac{1}{1-L}\int_0^\infty e^{\alpha t}dF(t)\right\}\\&=\frac{1}{\alpha}\left\{-\frac{L}{1-L}+\frac{1}{1-L}\right\}=\frac{1}{\alpha},\end{aligned}$$

这里只要注意到由条件(23)有 $e^{\alpha t}[L-F(t)]\to 0,t\to\infty$. □

推论 1 设 F 为满足(17)的非格子点分布,条件(23)满足,则

$$\lim_{t\to\infty}e^{\alpha t}[m(t)-m(\infty)]=-\frac{1}{\alpha}\left[\int_0^\infty te^{\alpha t}dF(t)\right]^{-1}.\tag{27}$$

证 更新函数 $m(t)$满足更新方程(22),其中 $h(t)=F(t)$.因此 $h(\infty)=L$,且

$$h_1(t)=\frac{L-F(t)}{1-L},\qquad \int_0^\infty \mathrm{e}^{\alpha t}h_1(t)\mathrm{d}t=-\frac{1}{\alpha}.$$

由此即得(25). □

推论 2 设 F 为满足(15)的非格子点分布,条件(21)满足,则

$$\lim_{t\to\infty}\mathrm{e}^{\alpha t}\mathrm{P}(\zeta>t)=\frac{1-L}{\alpha\hat{\mu}}. \tag{28}$$

证 由(20)式可验证 $z(t)=\mathrm{P}(\zeta\leqslant t)$满足更新方程

$$z(t)=(1-L)+\int_0^t z(t-s)\mathrm{d}F(s). \tag{29}$$

现在利用定理 4.6.这时 $z(\infty)=1,z(\infty)-z(t)=\mathrm{P}(\zeta>t),h(t)=1-L=h(\infty)$,由(23)即得(26). □

例 4.4 保险破产问题 设 N 是参数为 α 的泊松过程,$\{Y_n,n\geqslant1\}$为 i.i.d.随机变量,其共同分布为连续(非格子点)分布 $G,G(0)=0,\mu=\int_0^\infty x\mathrm{d}G(x)<\infty$,且$\{Y_n,n\geqslant1\}$与 N 独立.

我们把 N 的第 n 个更新时刻 S_n 解释为某保险公司的第 n 个要求赔偿的顾客来到的时刻,Y_n 为这个顾客得到的赔偿金额.假定这家保险公司的初始资本为 $x>0$,收益率(单位时间的收入)为 $c>0$,则时刻 t 公司的资产为

$$V(t)=x+ct-\sum_{n=1}^{N(t)}Y_n,\qquad t\geqslant0,$$

公司永远不破产的概率为

$$B(x)=\mathrm{P}\Big(\forall\, t>0,x+ct-\sum_{n=1}^{N(t)}Y_n\geqslant0\Big). \tag{30}$$

约定 $x<0$ 时,$B(x)=0$,则

$$\mathrm{P}(\forall\, t>0,V(t)\geqslant0\mid W_1=s,Y_1=y)=B(x+cs-y),$$

这里 W_1 是 N 的第一个更新时刻,从而按全概率公式得

$$B(x)=\int_0^\infty\int_0^{x+cs}B(x+cs-y)\lambda\mathrm{e}^{\lambda s}\mathrm{d}s\mathrm{d}G(y).$$

作变换 $x+cs=t$ 可得

$$\mathrm{e}^{-\lambda x/c}B(x)=\frac{\lambda}{c}\int_0^\infty\left(\int_0^t B(t-y)\mathrm{d}G(y)\right)\mathrm{e}^{-\lambda t/c}\mathrm{d}t.$$

由此可知 $B(x)$是可求导的,将上式两边求导得

$$\mathrm{e}^{-\lambda x/c}\left[B'(x)-\frac{\lambda}{c}B(x)\right]=-\frac{\lambda}{c}\mathrm{e}^{-\lambda x/c}\int_0^x B(x-y)\mathrm{d}G(y),$$

$$B'(x)=\frac{\lambda}{c}B(x)-\frac{\lambda}{c}\int_0^x B(x-y)\mathrm{d}G(y),$$

从而

$$B(x)-B(0)=\frac{\lambda}{c}\int_0^x B(z)\mathrm{d}z-\frac{\lambda}{c}\int_0^x\left(\int_0^z B(z-y)\mathrm{d}G(y)\right)\mathrm{d}z,$$

$$\int_0^x\left(\int_0^z B(z-y)\mathrm{d}G(y)\right)\mathrm{d}z=\int_0^x\left(\int_y^x B(z-y)\mathrm{d}z\right)\mathrm{d}G(y)$$

$$=\int_0^x\int_0^{x-y}B(u)\mathrm{d}u\mathrm{d}G(y)=\int_0^x B(u)\mathrm{d}u-\int_0^x[1-G(y)]B(x-y)\mathrm{d}y.$$

所以

$$B(x)=B(0)+\frac{\lambda}{c}\int_0^x B(x-y)[1-G(y)]\mathrm{d}y.$$

由于 $\mu=\int_0^\infty[1-G(y)]\mathrm{d}y<\infty$，则

$$B(x)=B(0)+\frac{\lambda\mu}{c}\int_0^x B(x-y)\mathrm{d}G_e(y). \tag{31}$$

1) 设 $\frac{\lambda\mu}{c}<1$. 令 $F(y)=\frac{\lambda\mu}{c}G_e(y)$，则 $L=F(\infty)=\frac{\lambda\mu}{c}<1$，

$$B(x)=B(0)+\int_0^x B(x-y)\mathrm{d}F(y),$$

$$B(x)=B(0)[1+m(x)]. \tag{32}$$

按定义(30)，我们有 $\lim\limits_{x\to\infty}B(x)=1$.（事实上，由强大数律，以概率 1

$$\frac{1}{t}\left(ct-\sum_{n=1}^{N(t)}Y_n\right)\to c-\lambda\mu>0,$$

因此 t 充分大时 $ct-\sum_{n=1}^{N(t)}Y_n>0$，而在任何有界区间上 $ct-\sum_{n=1}^{N(t)}Y_n$ 有下界，故以概率 1 $\inf_{t\geqslant0}\left(ct-\sum_{n=1}^{N(t)}Y_n\right)>-\infty$.）再由(32)式可得

$$1=B(0)[1+m(\infty)]=B(0)\frac{1}{1-L},$$

$$B(0)=1-L=1-\frac{\lambda\mu}{c}. \tag{33}$$

又设有正数 α，使得

$$\int_0^\infty \mathrm{e}^{\alpha x}\mathrm{d}G_e(x)=\frac{c}{\lambda\mu},$$

即

$$\int_0^\infty \mathrm{e}^{\alpha x}[1-G(x)]\mathrm{d}x=\frac{c}{\lambda},$$

及

$$\int_0^\infty x e^{\alpha x}[1-G(x)]dx<\infty,$$

则由定理4.6，

$$\lim_{x\to\infty} e^{\alpha x}[1-B(x)]=\left(1-\frac{\lambda\mu}{c}\right)\left\{\frac{\alpha\lambda}{c}\int_0^\infty x e^{\alpha x}[1-G(x)]dx\right\}^{-1}. \tag{34}$$

(34)表明 $1-B(x)$ 按指数 $e^{-\alpha x}$ 减少，依此公式可确定 x 使破产概率 $1-B(x)$ 充分地小. α 称为伦贝格(Lundberg)系数.

2) 若 $\frac{\lambda\mu}{c}=1$，则(31)变成

$$B(x)=B(0)+\int_0^x B(x-y)dG_e(y),$$

从而

$$B(x)=B(0)[1+m_{G_e}(x)],$$

其中 m_{G_e} 是以 G_e 为等待时间分布的更新过程的更新函数，但

$$\lim_{x\to\infty} m_{G_e}(x)=\infty,\qquad 0\leqslant B(x)\leqslant 1,$$

故只能 $B(0)=0$. 所以 $B(x)\equiv 0$，即保险公司以概率1要破产.

3) 若 $\frac{\lambda\mu}{c}>1$，则存在唯一的 $\alpha>0$ 使得

$$\int_0^\infty e^{-\alpha x}dG_e(x)=\frac{c}{\lambda\mu}.$$

记

$$\hat{G}_e(x)=\frac{\lambda\mu}{c}\int_0^x e^{-\alpha y}dG_e(y),\qquad \hat{B}(x)=e^{-\alpha x}B(x),$$

则

$$\hat{B}(x)=B(0)e^{-\alpha x}+\int_0^x \hat{B}(x-y)d\hat{G}_e(y).$$

由关键更新定理，

$$\lim_{x\to\infty} e^{-\alpha x}B(x)=B(0)\left\{\frac{\alpha\lambda\mu}{c}\int_0^\infty x e^{-\alpha x}dG_e(x)\right\}^{-1}.$$

然而由于 $B(x)$ 有界，$\lim\limits_{x\to\infty} e^{-\alpha x}B(x)=0$，所以 $B(0)=0$，从而 $B(x)\equiv 0$，保险公司同样以概率1要破产. □

例4.5 分支过程 设某种生物的寿命分布为 F，F 为非格子点分布. 每个生物在它的生命结束时，以概率 p_j 产生 j 个后代，$j=0,1,\cdots$. 设

$$\gamma = \sum_{j=1}^{\infty} jp_j < \infty .$$

各个生物的寿命及产生后代的个数均相互独立,以 $X(t)$记时刻 t 活着的生物个数.设 $X_0=1$.记

$$m(t) = \mathrm{E}[X(t)].$$

初始生物的寿命记为 T,则

$$\mathrm{E}[X(t) \mid T = s] = \begin{cases} 1, & s > t, \\ \gamma m(t-s), & s \leqslant t, \end{cases}$$

因为 $s>t$ 时,$X(t)=1$,而 $s\leqslant t$ 时,以 s 为时间原点,初始生物的每个后代到时刻 t 平均产生 $m(t-s)$个后代,而初始生物平均有 γ 个后代,所以

$$m(t) = \mathrm{E}[X(t)] = \int_0^{\infty} \mathrm{E}[X(t) \mid T = s] \mathrm{d}F(s) = \int_t^{\infty} \mathrm{d}F(s) + \int_0^t \gamma m(t-s) \mathrm{d}F(s),$$

$$m(t) = \overline{F}(t) + \gamma \int_0^t m(t-s) \mathrm{d}F(s). \tag{35}$$

1) 假设 $\gamma>1$,这时存在唯一的正数 α,使得

$$\int_0^{\infty} \mathrm{e}^{-\alpha t} \mathrm{d}F(t) = \frac{1}{\gamma}.$$

定义

$$G(t) = \gamma \int_0^t \mathrm{e}^{-\alpha s} \mathrm{d}F(s), \qquad \hat{m}(t) = \mathrm{e}^{-\alpha t} m(t).$$

由(35),

$$\hat{m}(t) = \mathrm{e}^{-\alpha t} \overline{F}(t) + \int_0^t \hat{m}(t-s) \mathrm{d}G(s). \tag{36}$$

$G(t)$是概率分布函数:$G(\infty)=1$,且

$$\int_0^{\infty} \mathrm{e}^{-\alpha t} \overline{F}(t) \mathrm{d}t = \int_0^{\infty} \mathrm{e}^{-\alpha t} \int_t^{\infty} \mathrm{d}F(s) \mathrm{d}t = \int_0^{\infty} \int_0^s \mathrm{e}^{-\alpha t} \mathrm{d}t \mathrm{d}F(s)$$
$$= \frac{1}{\alpha} \int_0^{\infty} (1 - \mathrm{e}^{-\alpha s}) \mathrm{d}F(s) = \frac{1}{\alpha}\left(1 - \frac{1}{\gamma}\right).$$

由更新方程(33)及关键更新定理,

$$\lim_{t \to \infty} \mathrm{e}^{-\alpha t} m(t) = \frac{\gamma - 1}{\alpha \gamma^2 \int_0^{\infty} t \mathrm{e}^{-\alpha t} \mathrm{d}F(t)}.$$

这表明平均人口按指数率增长,α 称为马尔萨斯(Malthus)增长率.

2) 若 $\gamma=1$,$\mu = \int_0^{\infty} t \mathrm{d}F(t) = \int_0^{\infty} \overline{F}(t) \mathrm{d}t < \infty$,由(35)及关键更新定理

$$\lim_{t \to \infty} m(t) = \frac{1}{\mu} \int_0^{\infty} \overline{F}(t) \mathrm{d}t = 1.$$

3) 若 $\gamma<1$,则

$$m(t)=\overline{F}(t)+\sum_{n=1}^{\infty}\gamma^{n}F^{*n}*\overline{F}(t).$$

由于$\lim\limits_{t\to\infty}\overline{F}(t)=0,\lim\limits_{t\to\infty}F^{*n}*\overline{F}(t)=0$,因此

$$\lim_{t\to\infty}m(t)=0.$$

若又存在正数 $\alpha>0$,使得

$$\int_0^{\infty}\mathrm{e}^{\alpha t}\mathrm{d}F(t)=\frac{1}{\gamma},$$

则由定理4.6,

$$\lim_{t\to\infty}\mathrm{e}^{\alpha t}m(t)=-\frac{1-\gamma}{\alpha\gamma^{2}\int_0^{\infty}t\mathrm{e}^{\alpha t}\mathrm{d}F(t)}.\qquad\square$$

习　　题

4-1. 设在有偿更新过程$\{X(t),t\geqslant0\}$中,W_n 与 Y_n 独立,$\mathsf{E}[|Y_1|]<\infty$,则

$$\mathsf{E}[X(t)]=\mathsf{E}[Y_1]\mathsf{E}[N(t)].$$

4-2. 设更新过程 N 的等待时间分布F 满足条件$q=F(0)>0$.记

$$\widetilde{F}(t)=\begin{cases}\dfrac{F(t)-q}{1-q}, & t\geqslant0,\\ 0, & t<0.\end{cases}$$

设更新过程 $\widetilde{N}=\{\widetilde{N}(t),t\geqslant0\}$以 $\widetilde{F}$ 为等待时间分布,$\{Y_n,n\geqslant0\}$为 i.i.d.随机变量序列,与$\widetilde{N}$相互独立,且

$$\mathsf{P}(Y_n=k)=(1-q)q^{k-1},\qquad k\geqslant1.$$

令

$$X=\left\{X(t)=(Y_0-1)+\sum_{n=1}^{\widetilde{N}(t)}Y_n,t\geqslant0\right\},$$

则 X 与N 同分布.

4-3. 设 $h(t)$为局部有界函数,且 $\lim_{t\to\infty}h(t)=h<\infty$,$z(t)$是更新方程

$$z(t)=h(t)+\int_0^{t}z(t-s)\mathrm{d}F(s)$$

的解,则

$$\lim_{t\to\infty}\frac{z(t)}{t}=\frac{h}{\mu}.$$

4－4. 设 $z(t)$是更新方程

$$z(t)=t^{n-1}+\int_0^t z(t-s)\mathrm{d}F(s)$$

的解，求$\lim_{t\to\infty}[z(t)/t^n]$.

4－5. 来到某汽车站的旅客服从强度为 λ 的泊松过程. 当车站上有 L 个旅客时便由一辆汽车将旅客运走. 现假定，当车站上有 n 个旅客等待时，车站每单位时间将支出服务费用 nc 元，每开出一辆汽车支出费用 K 元. 试问长时间后车站每单位时间的平均支出费用为多少？这个量显然对确定汽车的票价有重要作用.

4－6. 设一部售米机的储存箱可装 S 公斤大米. 顾客按一更新过程来到售米机前购米，顾客来到间隔服从非格子点分布 F. 各个顾客的购米量服从参数为 λ 的指数分布，且相互独立. 假定售米机中储存的大米少于 s 公斤时，将会在瞬间重新装满. 以 $X(t)$记时刻 t 售米机中的大米公斤数，$s<x<S$，证明：以概率 1

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t 1_{\{X(s)\geqslant x\}}\mathrm{d}s=\lim_{t\to\infty}\mathrm{P}(X(t)\geqslant x)=\frac{1+\lambda(S-x)}{1+\lambda(S-s)}.$$

4－7. 对习题 3－11 中定义的多状态交替更新过程$\{V(t),t\geqslant 0\}$证明：以概率 1

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t 1_{\{V(s)=j-1\}}\mathrm{d}s=\frac{\mathrm{E}[X_1^{(j)}]}{\mathrm{E}[X_1^{(1)}+\cdots+X_1^{(d)}]},\quad j=1,\cdots,d.$$

4－8. 甲、乙、丙三名射手轮流射击一个靶子. 射击规则是射中者应再射，直至首次射不中为止. 以此规则不断地轮流射击. 三名射手每次射中的概率分别为 p_i，$i=1,2,3$，并且假设各次射击的情况相互独立. 试问：经过长时间后，甲、乙、丙的射击次数的比例是多少？

4－9. 设乘客按强度为 λ 的泊松过程来到火车站乘车，火车按一更新过程来到车站载客，每列火车将候车的乘客全部载走，火车来到间隔服从非格子点分布 F，且 $\mu_F<\infty$. 以 $X(t)$记时刻 t 在候车的乘客数. 证明：以概率 1

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t 1_{\{X(s)=i\}}\mathrm{d}s=\lim_{t\to\infty}\mathrm{P}(X(t)=i)=\frac{1}{\mu_F}\int_0^\infty\sum_{j=i}^\infty\frac{\lambda^j t^{j+1}}{(j+1)!}\mathrm{e}^{-\lambda t}\mathrm{d}F(t),\quad i\geqslant 0.$$

参考书目

[1] 费勒 W. 概率论及其应用(下册).刘文译.北京:科学出版社,1979
[2] 钟开莱.初等概率论附随机过程.魏宗舒等译.北京:人民教育出版社,1980
[3] 复旦大学.随机过程.北京:人民教育出版社,1981
[4] 侯振挺.Q-过程的唯一性准则.长沙:湖南科技出版社,1982
[5] 胡迪鹤.应用随机过程引论.哈尔滨:哈尔滨工业大学出版社,1984
[6] 陆大铨.随机过程及其应用.北京:清华大学出版社,1986
[7] 帕尔逊 E. 随机过程.邓永录,杨振明译.北京:高等教育出版社,1987
[8] 汪荣鑫.随机过程.西安:西安交通大学出版社,1987
[9] 闵华玲.随机过程.上海:同济大学出版社,1987
[10] 吴俊杰,潘麟生.随机过程.长沙:湖南大学出版社,1988
[11] 何声武.随机过程导论.上海:华东师范大学出版社,1989
[12] 申鼎煊.随机过程.武汉:华中理工大学出版社,1990
[13] 施仁杰.马尔科夫链基础及其应用.西安:西安电子科技大学出版社,1992
[14] 邓集贤,许刘俊.随机过程.北京:高等教育出版社,1992
[15] 方兆本,缪柏其.随机过程.合肥:中国科技大学出版社,1993
[16] 何镇邦,李桂荣.概率论与数理统计附随机过程.北京:兵器工业出版社,1993
[17] 郭绍建,萧亮壮,张福渊,付丽华.概率统计及随机过程.北京:航空工业出版社,1993
[18] 陈永义.马尔科夫链——理论、应用与算法.兰州:兰州大学出版社,1993
[19] 严颖,成世学,程侃.运筹学随机模型.北京:中国人民大学出版社,1995
[20] 邓永录.随机模型及其应用.北京:高等教育出版社,1994
[21] 许升汉.概率论与随机过程.北京:人民邮电出版社,1996
[22] 陈木法.跳过程与粒子系统.北京:北京师范大学出版社,1986